THE EFFECTS OF FEEDING REGIME ON LAMB FLAVOUR

Doreen A. Parry

Doctor of Philosophy (Ph.D)

University of Edinburgh

1983

DECLARATION

I declare that the experimental work described in this thesis is entirely my own. As indicated, Mrs M P Woods collaborated in the Pilot Study in the section Related Studies but from that time onwards was not involved in any of the subsequent experimental programmes except as a taste panel member. I acknowledge valuable discussions relating to experimental design and statistical techniques with Miss Elspeth M Morrison, Queen Margaret College and with Mr M M Franklin, Statistician, ARC.

1983

UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS

Name of Candidate Mrs Doreen Anna Parry

Address

Degree PhD *Date* 1983

Title of Thesis The Effect of Feeding Regime on Lamb Flavour.

Following reports that grazing lucerne and rape had produced flavour defects in lamb, the aim of this study was to determine if any of the fattening regimes for store lambs in the ESCA trials of 1977 produced unacceptable aroma or flavour in roasted gigots and loins. Careful studies of factors influencing meat flavour were made so that standardised procedures would ensure that feeding regime was the only variable in the trials. Hedonic rating scales were first used to determine optimal internal temperature for gigots and loins. Tasters at QMC and ESCA found difficulty in distinguishing samples cooked to different internal temperatures. Responses were inconsistent in duplicated trials. It was thus considered prudent to devise more precise assessment techniques for future studies. Triangle tests and paired comparison (preference) tests were subsequently used. In the first comparisons of grass and rape fed samples, there were no detectable differences in aroma or flavour. There was no consistent preference for either sample. Flavour profiles confirmed that neither feeding regime produced characteristic aroma or flavour. Tasters' performance was assessed. Results of these assessments indicated that the second series of comparisons of grass and rape and comparisons of the other forage crops could be carried out with greater confidence in both the experimental design and tasters' performance. None of the fattening regimes for store lambs produced detectable differences in aroma or flavour. There was no consistent preference for the flavour of lamb from any of the feeding regimes although there was some evidence that feeding cabbage produces flavour which is very well liked. There was no difference in the ability of judges to identify the odd sample in triangle tests when the aroma of raw and cooked samples was compared. This finding has considerable implications for future experiments. The contribution of fatty tissue to flavour was studied. Total and evaporative weight losses and raw and cooked pH values of gigots and loins were recorded. Statistically significant differences were established between the two cuts although pH changes in response to the cooking process did not differ. These findings emphasise the importance of comparing meats of standard anatomical location. An account of the basis on which the sensory tests used in the trials were selected is given together with details of statistical techniques used in the analysis of data.

CONTENTS

ABBREVIATIONS USED

AAHE	American Association of Home Economists
ARC	Agricultural Research Council
ATP	Adenosine Triphosphate
BS	British Standard
CSIRO	Commonwealth Scientific and Industrial Research Organisation
EEC	European Economic Community
ESCA	The Edinburgh School of Agriculture
IFST	Institute of Food Science and Technology
MRI	Meat Research Institute, Bristol, U.K.
MRL	Meat Research Laboratory, Cannon Hill, Queensland, Australia
QMC	Queen Margaret College, Edinburgh
SCI	Society of Chemical Industry
SD	Standard Deviation
$\bar{X}$	Arithmetic Mean

CHAPTER 1

Introduction

1. Lamb and Human Nutrition

The British are the third largest eaters of lamb within the EEC and produce approximately fifty per cent of the sheep within the EEC. Approximately 57 per cent of lamb consumed is produced within the United Kingdom with 40 per cent being imported from New Zealand (Prodfact 1982). Lamb is thus a significant flesh food within the United Kingdom. Comparative figures for the consumption of mutton* and lamb and other carcass meats are given in Table 1.1.

Table 1.1 Household Consumption of Carcass Meats 1979

(oz. per person per week)

	United Kingdom	(g)	Scotland	(g)
Beef and Veal	8.27	235	11.57	329
Mutton and Lamb	4.28	126	1.86	53
Pork	3.63	103	1.79	51
Total	16.18	464	15.22	33

Source: National Food Survey 1979

(Figures quoted in oz. have been converted to g for clarity).

The differing patterns of consumption for Scotland and the United Kingdom as a whole should be noted. In Scotland, contrary to popular opinion, consumption of meat products was only 23.45 oz (666g) per person weekly compared with the United Kingdom as a whole where average consumption was 124.09 oz (684g). In making these assessments the conventional value of ten per cent has been deducted for wastage. Allowance was made for inedible matter. Although six per cent edible food was reported as being either discarded or fed to pets (Wenlock et al 1980) this figure was considered by the National Food Survey Committee to be a conservative estimate.

In/

*Mutton is produced from large carcasses and a small amount of ewe mutton is produced mainly for specialised restaurant and ethnic requirements.

In Table 28 of the National Food Survey (1979) it is suggested that at a time when deep freeze ownership is increasing and meats are commonly bought in bulk frozen, United Kingdom figures should perhaps be modified to take these factors into account (Table 1.2).

Table 1.2 Household Consumption of Carcass Meats 1979

oz. per person per week

	Freezer Owners	(g)	Others	(g)	Total	(g)
Beef and Veal	9.06	257	7.59	216	8.27	235
Mutton and Lamb	4.69	133	3.91	111	4.28	126
Pork	4.10	116	3.22	91	3.63	103
Total	17.85	506	14.72	418	16.18	464

Figures for different regions are not quoted separately in the tables. Freezer owners certainly purchase more lamb but it does not follow that more is actually consumed.

When nutrient intakes of the United Kingdom as a whole and Scotland are compared, (Appendix Table 1.1) variations are slight. Hence it seems appropriate to consider the proportions of the daily energy, protein, fat and iron intakes derived from carcass meats to be reasonably representative of those for Scotland. These are set out in Appendix Table 1.2. Despite lower lamb and pork consumption in Scotland gross compositional differences of meats of the three species will be slight in comparison with variability of individual meat samples. Percentages of total daily intake of nutrients provided by carcass meats are set out in Table 1.3.

Table 1.3/

Table 1.3 Percentages of Total Daily Intake of Nutrients Provided by Carcass Meats

Nutrient	% Contribution to Daily Intake
Energy	6.6
Protein	13.9
Fat	11.6
Iron	8.5
Saturated fatty acids	11.1
Polyunsaturated fatty acids	6.0

At a time when the James Report compiled by the National Advisory Committee on Nutrition Education is still under discussion, the relatively high contribution of carcass meats to protein and to a lesser extent fat intake compared with their low energy contribution should be noted. It can be concluded that meats remain a good source of protein of high biological value. Experiments designed to improve and make meat production more efficient without sacrifice of eating quality are thus of obvious value.

2. Sensory Testing of Foods Produced on ESCA Farms at Queen Margaret College

When the author joined the staff of QMC in 1964, then the Edinburgh College of Domestic Science, links had already been established with ESCA. Lectures from ESCA staff and farm visits were arranged at that time. Assessments of new varieties of potato cooked by different methods were carried out. QMC staff assisted in the comparisons of beef samples from different breeds which had been cooked under standard conditions (Vandore 1968). A study of the effects of breed, housing conditions and diet on the eating quality of hens' eggs had been initiated.

During the academic session 1966-67, the author was attempting to ensure/

ensure that assessments of appearance, flavour and texture of foods by students of Home Economics and Catering was made on a sound, scientific basis. Discussions with Miss M P Dixon* of Arthur D Little Research, Inveresk, a highly trained assessor of international reputation, combined with student visits to the Institute aroused great interest in developing new techniques. Detailed studies of the sensory mechanisms involved in the assessment of foods which were combined with a study of the uses and limitations of sensory (organoleptic) appraisal techniques proved to be of absorbing interest. It is hoped that the author's enthusiasm and interest were evident from a paper which was presented at the Sixteenth Easter School in Agricultural Science, University of Nottingham in 1969 (Parry 1970).

Records of students' performance in detecting primary tastes, ranking sucrose solutions and odour recognition were started in 1968 and have continued until the present time. Hedonic rating scales were used to assess a wide variety of foods and to study the effects of flavour modifiers and enhancers. Sensory appraisal techniques became more formalised following an In-Service course for practising teachers. The amount of data collected had by them become considerable.

The Data Processing Centre was established at QMC in 1976. There was a telephone link with a main frame computer. It was thus hoped to analyse the sensory appraisal data with computer support. To do so, considerable discussion of the design of experiments and methods of statistical analysis of results obtained was required. In particular, the Systems Analyst had considerable reservations about treating data derived from the hedonic rating scales as other than ordinal.

Methods/

*now Mrs M P Woods, Lecturer, QMC.

Methods for the analysis of data from paired comparisons and triangle tests were however already well established (Amerine et al 1965). Thus from 1976 onwards, the design of experiments was discussed with the Systems Analyst and analysis of results was made with computer support. The experimental design of the present study was discussed and agreed with ARC's statistician, M M Franklin.

During 1975, the first formal joint ESCA/QMC experimental programmes started. An account of a pilot study to determine if cross-contamination occurred when meats to be compared were roasted on the same shelf of a preheated oven and of the feasibility of using triangle tests to detect differences in the aroma of fat from pigs fed a diet including deoiled herring silage from fat from matched controls is included in the section Related Experimental Studies. Also included are a comparison of bull versus steer meats and tests carried out on growing pigs to assess the effects of the inclusion of deoiled herring silage in the diet. Paired Comparisons and Triangle Tests on flavour were carried out with the assistance of final year student Home Economists in 1976. An account of the flavour of pork from pigs fed diets containing grain distillers' spent wash carried out at QMC in May 1977 is also included. This work was not published but was reported in the Annual Report of the East of Scotland College of Agriculture of 1979.

3. Aims and Objectives of the Present Study

In 1977, a long-term project was set up on ESCA farms to compare the performance of store lambs finished on a variety of fattening systems. Scottish Blackface wether lambs, born in April/May 1977, after weaning in August were grazed on grass at lower altitudes until allocation to their selected crops. Groups were randomised by weight and/

and condition at the start of the experiment. The design of this experiment is indicated in Appendix Table 9.3. Lambs were weighed at three weekly intervals and sold fat as they reached marketable condition by subjective assessment. Carcass weights and grades were recorded at slaughter. Crop yields were recorded prior to and after grazing. As indicated in the Annual Report of the East of Scotland College of Agriculture of 1978, representative samples of these store lambs were used for consumer appraisal. These were the lambs used for most of the trials described in this study including the comparisons of the grass silage and lucerne silage fed lambs. The second series of comparisons between grass and rape fed lambs used April/May born lambs from the 1978 replicate. The same cultivar of rape (Lair) was used in both years.

Prior to 1977, there had been reports of adverse effects of grazing crops such as clover, lucerne and certain varieties of brassica and in particular rape, on lamb flavour.* Since both rape and lucerne silage had been used in the 1977 ESCA trials as well as other varieties of brassica, it was considered desirable to establish if any of the crops used in the 1977 trials were likely to cause similar adverse effects on lamb flavour. This had become possible as a result of expertise in aroma and flavour assessments of meats acquired by staff and students of QMC during the assessments of bull and steer meats and pork which had been carried out since 1975.

In these studies, interest had been focussed on identifying detectable differences in aroma and flavour by the use of triangle tests combined with determining flavour preferences as assessed by paired comparison tests.

The/

*References are given at the end of this chapter.

The main aims of the trials described in the present study were to establish if detectable differences in the aroma and flavour or significant flavour preferences were produced as a result of any of the feeding regimes used in the ESCA trials. If detectable aroma or flavour differences were present, it was also important to determine if these differences were linked in any way to preference. Triangle tests were used to determine if differences in aroma or flavour did or did not exist between samples. Preferences were established by the use of paired comparisons. This technique was selected in view of reservations of the Systems Analyst concerning some methods of statistical analysis of data derived from hedonic rating scales and because subjects at QMC were so erratic and inconsistent in their performance as to make it unlikely that valid results would be obtained using this technique.

It was recognised that in using such procedures that if statistically significant differences in aroma or flavour were produced by a particular feeding regime, further tests of a different type, likely to be descriptive in nature, would be required. Thus the first priority was to demonstrate if detectable differences were present. Only if this were so could further experimentation be justified.

Many of the previous studies of the effect of feeding regime on lamb flavour were based on work pioneered by Cramer et al (1967). Many laboratories where assessments of lamb flavour are made use hedonic rating scales of various types, often following a series of difference tests.

Assessment of aroma and flavour are based on scores linked to favourable and unfavourable characteristics. In this respect such results are not as precise as the results of difference tests used in the present study.

4. Presentation and Development of the Present Study

A study of factors influencing meat flavour (Chapter 2) emphasises the importance of standardising procedures so that the only variable under study is that of feeding regime. For this reason, lambs remained on their respective crops until slaughter.

The basis of the sensory appraisal techniques and the principles of the statistical techniques used in the analysis of data are indicated in Chapter 3.

Although details of the methods of preparation of samples for presentation to tasters in other laboratories were available, it was considered important to devise a technique which would relate closely to United Kingdom consumer practice whilst at the same time ensuring control over variables. The procedures which were followed and the basis of their selection are discussed in Chapter 4.

Before the main trials, two preliminary studies were made to determine optimal internal temperature for gigots and loins roasted at a temperature of 177°C. This temperature is 14°C higher than that used in many other laboratories. Four different internal temperatures were selected to ensure that samples were prepared considered by participants to be acceptable. Hedonic rating scales were used to establish tasters' attitude to each batch of samples (Chapter 5). From results of these tests it became evident that responses of tasters in hedonic rating scales showed such variation and lack of consistency that it was considered prudent to devise an alternative form of experimental design for the main trials.

Thus for the first Grass versus Rape trials, (Chapter 6) it was decided to concentrate on determining if differences in aroma and flavour and consistent preference for lamb from either feeding regime could be demonstrated. Results indicated that no statistically

significant/

significant differences or preferences were established in these trials. This could have been a result of either poor performance by tasters or because differences between samples, if present, were so small as to be undetectable.

In view of results described in Chapters 5 and 6, it was considered important to assess subjects' acuity and to confirm that the methodology devised was appropriate to the aims and objectives of the study. Thus two very different lamb samples were compared. Young ('spring') lamb from barley fed females housed indoors and slaughtered at four to five months was compared with much older turnip fed wethers reared outdoors and slaughtered at approximately ten months. Results of these trials are described in Chapter 7. In addition a flavour profile analysis was made on gigots and loins from the first batch of grass and rape fed samples. Flavour profiles on three samples - two rape and one grass fed - indicated that three different samples were being assessed rather than two from one feeding regime and one from another in both gigots and loins. Thus it was not perhaps surprising that differences had not been established between samples in trials described in Chapter 6. Results of trials described in Chapter 7 indicated that differences in aroma and flavour between the two samples were readily established. Preferences (of statistical significance) were for the older lamb. Preference is highly subjective and it is to be expected that consistent preference for one sample of a pair will be demonstrated only if one of them is either extremely pleasant or unpleasant in comparison to the other. Hence tasters' acuity and their responses were vindicated by this series of trials.

Following the trials described in Chapter 7, it was possible to repeat the Grass versus Rape trials (Chapter 8) and studies of other forage crops (Chapter 9) with greater confidence in both the experimental/

experimental design and tasters' performance. The experimental design therefore remained unchanged throughout subsequent trials. The only exception was that aroma assessments were discontinued during the second Grass versus Rape trials. Subjects had found them so extremely distasteful that they became reluctant to continue with the trials. They also sapped confidence. Because there seemed to be greater differences between grass and rape fed samples in the second series of trials, it was fortunate that aroma assessments were made at a later stage by BSc students of Agriculture at ESCA.

In Chapter 10, total and evaporative weights losses and pH values of raw and cooked gigots and loins are compared. Statistically significant differences are demonstrated between the two cuts.

Following results of the deoiled herring silage trials, it was considered that differences in aroma could be demonstrated more readily in raw than in cooked samples. Results of these tests are given in Chapter 11 where experiments to determine the significance of the contribution of fatty tissue to meat flavour are also described.

In Chapter 12, results of the present studies are discussed with reference to methodology, choice of sensory appraisal technique and statistical analysis of data in relation to studies carried out in other laboratories. Conclusions of importance to future experimental programmes and contributions to fundamental knowledge are highlighted.

5. Reports on the Effects of Feeding Regime on Lamb Flavour in Relation to the Present Study

Ford and Park (1980) give an account of a programme of research to investigate the flavour of sheep meat following complaints from consumers in North America and Japan. They considered that diet immediately prior to slaughter affects lamb flavour. Cramer et al (1967)/

(1967) and Shorland (1970) reported that grazing white clover (Trifolium repens) produced lamb of more intense flavour than perennial ryegrass (Lolium perenne). Rhodes (1970) working with red clover (Trifolium pratene) considered that any slight differenceswere unlikely to be detected by the consumer. Rape (Brassica napus L) was reported to produce an objectionable aroma and flavour by Park et al (1972b) and by Wheeler et al (1974). Lucerne (Medicago sativa) has also been implicated in producing flavour defects by Nicol and Jagusch (1971) and Park et al (1972a and 1975). Park et al (1972b) compared the flavour of lambs grazed vetch, rape and oats (Vicia sativa, Brassica napus and Avena sativa). These three crops produced meats of different flavour from pasture fed and feedlot lambs. Park and Minson (1972) also reported the effects of tropical legumes on lamb flavour. Vesely and Hironaka (1976) indicated that lambs fed a cereal diet of oats and barley supplemented with hay produced lamb with flavour scores higher than if cereals alone were used. These reports will be discussed more fully in Chapter 12.

6. Protected Lipid Supplements

Monogastric animals deposit fats with a fatty acid composition similar to that of the diet. The fatty acid content of fat depots of ruminants is modified as a result of microbially induced hydrogenation and metabolism in the rumen. Fatty acid composition can however be controlled by the use of 'protected' lipid supplements such as formaldehyde treated sunflower or safflower/casein mixtures. The effect of feeding such supplements is to increase the linoleic acid concentration of the fat depots. From 1970 onwards many dietitians and nutritionists favoured a reduction in intake of saturated fatty acids combined with an increased intake of polyunsaturated fatty acids. Although such recommendations are at present under discussion, it could be argued that/

that the use of protected lipid supplements would be of benefit in human nutrition.

Experiments were however carried out by Park et al from 1974 onwards with the aim of modifying the characteristic cooked flavour of lamb and mutton in such a way as to make it more acceptable in North America and Japan. Ford et al (1975) indicated that taste panels preferred pasture fed lamb with its more intense meaty flavour and less 'unusual' aroma to protected sunflower seed supplement fed lamb. Park et al (1974) identified 4-hydroxydodec-cis-6 enoic acid lactone as an important aroma component of fatty tissue in sheep fed on a lipid supplement. Park and Ford (1975) reported that descriptions such as 'oily' appeared after two weeks supplementation whereas the description 'sweet' was given only after three weeks. The 'oily' criticism was not observed in raw meats and is thought to arise during the cooking process as 2, 4-decadienal is formed. Meat from sunflower seed based supplement was criticised less than that from safflower seed (Park et al 1976). The use of these supplements to modify mutton flavour is reported by Park et al (1978a) and a study of the 'sweet' flavour of lamb is reported by the same authors (1978b). The use of formaldehyde treated whole fat soya beans or soya flour to protect polyunsaturated fats from rumen degradation is reported by Ackerson et al (1976) and by Jagusch (1976) on formaldehyde treated ground sunflower seed. Other studies are reported by Wright et al (1974), Purchas (1975) and Garrett et al (1976). Ford and Park (1980) considered that provided the 'sweet' flavour could be eliminated, the oily flavour of lamb might be more readily acceptable to those who like the oiliness of pork and poultry. The Head of Science Division, Meat Industry Research Institute of New Zealand indicated that feeding of protected lipid/

lipid supplements had been discontinued in New Zealand (Personal communication 1982). Whilst it is appreciated that the information in this section was not of direct relevance to the present study no review of the effects of dietary regime would have been complete without it.

7. Summary

As a result of careful experimentation it has been possible to show that none of the crops used in the ESCA store lamb trials of 1977 caused differences in the aroma or flavour of roasted gigots and loins. It proved difficult to distinguish between test and control samples in almost all trials even in the second series comparing grass and rape. Whilst it is desirable that these findings should be confirmed in view of previous reports of some of the crops producing lamb of unacceptable flavour, the present findings suggest that the consumer will not reject lamb from any of these feeding regimes particularly as it is likely that it will be eaten with accompaniments in a social setting rather than by itself in the rather artificial setting of a sensory appraisal room.

CHAPTER 2

Summary

A brief account of the perception of flavour is given and some of the problems of studying the chemistry of flavour in a complex food system like roasted meat are considered. The contributions of pre- and post slaughter treatments of meat animals, the effects of cooking processes and genetic environmental factors which affect lamb flavour are described.

Background of this Study of the Effects of Feeding Regime on Lamb

1. Flavour

The flavour of foods is a complex blend of aroma, taste and a characteristic known as 'mouthfeel'. Flavour perception is complicated in that flavour anticipation is affected by the appearance of foods and by auditory stimuli. Detailed accounts of perception of aroma and taste are available in physiology and psychology texts, in Moncrieff (1951), Amerine et al (1965), Moncrieff (1970), Teranishi (1971), Harper (1972), Kare and Maller (1977), Le Magnen and Macleod (1977), Cagan and Kare (1981) and in journals such as The Chemical Senses (formerly Chemical Senses and Flavour). In the context of the present study only essential features will be considered.

Odour receptors are present in the nasal olfactory epithelium. They respond to many stimuli which are difficult if not impossible to classify. Threshold values vary and can be extremely low - 2.86×10^{-9}M and 1.12×10^{-9}M for n-butanol and butanoic (Butyric) acid respectively. Possibly because thresholds are so low and individuals vary in their responses to a given stimulus, there appears to be no general agreement as to more precise values. It is however accepted that/

that odorants of appropriate chemical configuration become bound to receptor sites in the olfactory epithelium thus stimulating transmission of nervous impulses to the olfactory bulb and the olfactory centre of the brain for interpretation. Aroma of foods is assessed before they reach the mouth, during chewing and at the inspiration of swallowing. Odour is the most important component of flavour. Blockage of the nasal receptors produces foods which the consumer describes as 'tasteless'.

Taste perception is limited to the taste buds of the tongue and possibly areas of the soft palate. There are four primary tastes - sweet, salt, sour and bitter. Sweetness is detected most readily at the tip of the tongue and bitterness at the back. Saltiness is detected at the sides of the tongue and sourness at the sides towards the back. Proteins have been isolated from the tastebuds of cows and pigs which can bind sweet and bitter tasting molecules. The taste buds are innervated by a branch of the seventh cranial (facial) nerve in the anterior two thirds of the tongue. The posterior of the tongue is innervated by the glossopharyngeal (ninth) cranial nerve. Common structural characteristics of bitter tasting compounds have been studied but the ABH theory described by Shallenburger (1971) has helped to explain the sweetness of such divergent compounds as D-amino acids, sugars, saccharin, aspartame and acesulfame potassium. Taste thresholds are higher than odour thresholds by values as high as 10^4.

'Mouthfeel' comprises food texture, kinesthetic responses, pain, temperature and chemoreception of sensations described as biting, metallic, mouthcoating and astringent. Texture can exert marked effects both on odour and taste perception influencing as it does the release of volatile compounds and solutes.

The/

The reader will thus appreciate that food flavour is not synonymous with taste with the former being far wider in its implications than the latter. In sensory appraisal work, it is a matter of regret that the terms are so often interchanged.

The chemical basis of food flavours has received extensive study. The development of modern sophisticated separation techniques has been of only limited value in a study of complex food systems such as meats. Extraction procedures may be incomplete or cause chemical changes, relative contribution of identified components is difficult to establish and the effects of masking, interaction and contrast cannot be simulated. Experimenters remain very dependent on human tasters.

In addition foods are not simple aqueous or oily systems. Flavour compounds are embedded in a heterogeneous matrix of lipids, proteins, carbohydrates, water and other compounds which both influences their activity and causes interactions which may affect flavour perception. King et al (1978) quote many examples ranging from simple lipid/water systems and refer to varying absorption of flavouring compounds by aqueously dispersed denatured proteins according to the presence or absence of solutes and emulsifiers. Whilst not of direct relevance to meat itself as opposed to meat as a component of a meal, the amylose-fraction of cooked starch can also absorb flavouring compounds.

Nursten (1977) classified foods into four groups according to the number of compounds required to simulate flavour. This grouping (Table 2.1) indicates the complexity of studies of the chemistry of meat flavour.

Table 2.1/

Table 2.1 Classification of Foods According to Volatile Components

Group	Food	C.I.	Contributory Compounds
1	Lemon	Citral	-
"	Almonds	Benzaldehyde	-
2	Butter	2, 3-butanedione	Ethanol, Dimethylsulphide
"	Boiled Potatoes	2-ethyl-3-methoxy-pyrazine	Methional
3	Roast Beef	A very large number of compounds required	
"	Bread	A very large number of compounds required	
4	Strawberry	Cannot be reproduced	
"	Chocolate	Cannot be reproduced	

where CI = character impact compound, i.e. a single substance on which the flavour of a food largely depends.

Nursten recognised that such a classification is controversial and over-simplified but considered that it draws attention to the problems of identifying the chemistry of meat flavour.

In 1972, Rhodes published The Chemistry of Meat Flavour (Meat Research Institute Memorandum No.24). He drew attention to the difficulty in identifying meat species if pieces of very lean tissue are cooked identically and are subsequently tasted or smelled blind. The same difficulty occurs if water extracts are heated. Fatty tissue when heated produced typical species aroma but only if water soluble components were not first removed. He quoted Park et al (1972a and 1972b), Shorland et al (1970) and his own work of 1971. In the first three studies, it is suggested that compounds present in forage feeds were deposited in lamb fat. This was not so in his own work. Further consideration was also given to the role of fat in meat flavour later in the Memorandum.

In 1978 Rhodes again drew attention to the water soluble components of fatty tissue regretting that Chang and Peterson (1977) had not considered their contribution to meat flavour. He also acknowledged

that /

that greater analytical resources in identifying components of meat flavour which are presented simultaneously at the olfactory centre have not explained the chemical basis of olfaction. He considered that meats are too variable and too much influenced by post slaughter ripening cooking procedures for flavour profile analysis to be as successful as in more simple food systems. Some of these problems are also considered by Patterson in Paul and Palmer (1972), Cole and Lawrie (1975), Lawrie (1979), Wasserman (1979) and more recently by Mottram (1983). The contribution of fat to meat flavour was investigated in the present studies and will be considered in more detail in Chapters 11 and 12.

2. The conversion of muscle to meat has important implications for eating quality. Differentiation of the terms muscle and meat has perhaps been insufficiently emphasised. Whilst defining meat as striated muscle tissue, it should be recognised that biochemical and biophysical changes initiated at slaughter cause considerable differences in chemistry and structure of the tissue. These changes exert a profound effect during the cooking process.

Despite death of the animal, tissue activity continues under local control. In muscle tissue, energy is required to maintain temperature and cellular integrity. At slaughter, blood circulation ceases. Aerobic tissue respiration by the cytochrome enzyme system continues for a very limited period as oxygen supply fails. Thus in vivo resynthysis of ATP from creatine phosphate which is itself resynthesised by the oxidation of glucose and fatty acids ceases. Myosin ATP-ase and sarcoplasmic ATP-ase activity lower ATP concentration with consequent release of inorganic phosphate. This stimulates anaerobic glycolysis until at a pH of 5.4 - 5.5 - the ultimate pH - enzyme activity is inhibited. ATP/

ATP disappears from the tissue. The onset of rigor is established coinciding with the formation of the actomyosin complex. The concentration of lactic acid produced as a result of glycolysis varies from 90 - 120/μmoles/g. Rate and extent of fall in pH varies with environmental temperature, species, type of muscle and pre-slaughter treatments. Glycogen may still be present in the tissue at the onset of rigor.

Fall in pH is accompanied by lower water binding capacity (WBC). Ultimate pH brings muscle proteins closer to their isoelectric point. Denaturation of sarcoplasmic proteins combined with precipitation on myofibrillar actomyosin also lowers WBC. Ammonia and phosphate are released from adenosine diphosphate with the formation of inosine 5' monophosphate. This nucleotide has flavour modifying and enhancing properties and may thus influence meat flavour as well as emphasising mouthfeel characteristics. Inosine 5' monophosphate may be further degraded to form hypoxanthine - itself a flavouring compound - and ribose, which, although present only in low concentrations can take part in Maillard and caramelisation reactions during the cooking process, thus affecting flavour.

Collagen and elastin are believed to be unaffected by conditioning. Whilst the myofibrillar proteins become denatured, the actomyosin complex remains although there is some detachment of the actin filaments at the Z band. Denatured proteins are more susceptible to hydrolysis by cathepsins, thus producing peptides and amino acids many of which confer flavour as well as taking part in Maillard reactions. In the presence of iron, lipoxidases may be active as are lipases. Glucose concentration increases as a result of α-amylase activity if glycogen remains in the tissue. Its possible role in production of flavour compounds is similar to that of ribose. Ammonia, hydrogen sulphide, acetaldehyde, acetone/

acetone, diacetyl and many other compounds associated with food flavour have also been isolated from ripened meat.

These changes are described in many studies and the reader is referred in particular to Lawrie (1966 and 1979).

3. In Chapter 4, some aspects of the cooking process on lamb flavour are considered. In particular, comparison was made of flavour differences in meats cooked by dry and moist methods. Patterson (1975) reviews studies of the chemical basis of meat flavour. Parry (1970) reported that raw meats have a faint aroma and an insipid salty and metallic flavour. Cooking, with its consequent changes in proteins, lipids and other compounds causes the development of typical meat flavours. The contribution of fatty tissue to flavour is considered in Chapter 11 of this study. It seems likely however that flavour precursors are present in both fibres and fatty tissue. Patterson (1975) and Lawrie (1979) describe experiments in which lyophilised beef slurry was heated at 100°C (10^{-5}mm Hg). Low boiling point volatiles such as ammonia, methylamine, hydrogen sulphide, methyl mercaptan, formaldehyde and acetaldehyde were isolated. Whilst disagreeable in higher concentrations, in traces they may be acceptable components of food flavour. Most will volatilise during the cooking process. The higher boiling point fraction included lactate, alcohols and aldehydes and had a pronounced meaty oroma. Maillard and caramelisation reactions have an important contribution particularly in dry methods of cookery. Carbonyls, heterocyclics such as pyrazines, furans and pyrolles and a variety of sulphur compounds and possible 5' ribonucleotides (present in ripened tissue) may combine to confer meat flavour.

4. Factors associated with the living animal which influence flavour are/

are well documented Paul and Palmer (1975), Patterson (1977), Sink (1979) and Ford and Park (1980). Hence they will be outlined only briefly in this study. Factors of major importance will be described under the headings genetic and environmental.

Genetic Factors: Species

As indicated in an earlier section of this chapter, although meat flavour is species dependent, most studies have implicated the fat rather than the lean component. Differences are likely to arise through genetic control of lipid composition and metabolism. Fatty acids, triglycerides and phospholipids or their derivatives may be of significance: Hornstein and Crowe (1964), Pearson et al (1973). Studies on the quantitative genetic aspects are few and results are variable.

Breed

There are few and sometimes conflicting reports of the effects of breed and sire on sheep meat flavour even from the same laboratories (Ford and Park 1980). Woodhams et al (1966) reported no influence of sire on lamb flavour. Cramer et al (1970) scored intensity of mutton flavour from Rambouillet, Targher and Columbia lambs. Objectionable flavour increased with fineness of wool. There was a statistically significant difference between breeds ($p<0.01$). However in a second experiment these workers could not repeat these results. This might be explained by feeding practice. In the second experiment lambs were fed together for 75 days preslaughter. In the first experiment this was not so. Diets could thus have differed. Vesely and Hironaka (1976) reported that neither breed nor sex had an effect on flavour scores. Dransfield et al (1978) demonstrated that sire and/

and breed had no influence on lamb flavour. Ford and Park (1980) report conflicting results of experiments on the effect of breed by Cramer et al (1970).

Sex.

Most reports on flavour refer either to pork or to differences in bull and steer meats. Ford et al (1972 unpublished) failed to establish differences in lamb flavour arising from either sex or breed. Wilson et al (1970) reported no difference in organoleptic traits between rams, cryptorchids and wethers. Corbett et al (1973) indicated that demonstrable flavour differences between chops from cryptorchids, wethers and females did not affect preference. Batcher et al (1969) indicated that broths prepared from some rams differ in flavour from wethers. Misock et al (1972) demonstrated higher scores of statistical significance from wethers as compared with rams of three different slaughter weights. However, Usborne (1961), Rhodes (1969 and 1970), Jeremiah et al (1972), Jacobs et al (1972), Vesely (1973) and Wong et al (1975) found no flavour differences attributable to sex. Methodologies varied between laboratories so it is difficult to compare these conflicting findings. However, it seems likely that the majority of consumers would not be influenced by the sex of sheep meat.

Environmental Factors: Age

As with beef, flavour intensity of lamb increases with age. This increase in flavour intensity parallels increasing myoglobin concentration and probably reflects changes in the metabolism of compounds such as amino acids, proteins, nucleotides and lipids. The latter appear to alter less. King et al (1970) reported more desirable lamb flavour in yearling and older than in young lambs. This was also noted in the present study (Chapter 7). Batcher (1979), Carpenter et al (1970) and Misock et al (1972) report that flavour scores decrease as flavour intensity increases with age. Carpenter et al (1970) noted

noted that the flavour of young lamb was bland and that the flavour of older animals was preferred. Paul et al (1964a) reported that yearling lamb had superior flavour to lamb of 5½ months at slaughter. Chemical studies of fatty tissue from lambs of different ages have been made by Wong (1972) who linked these to triangle tests by Sink and Caporaso (1977). Comments about practices in different laboratories in relation to effects of sex on flavour apply equally to studies of the effects of age on lamb flavour. With increasing age, meats become less tender and it is difficult to separate this characteristic completely from flavour.

Nutrition

The type of diet and nutrient source influence flavour. Whilst the amount and type of protein has little effect on lamb flavour there are reports that the amount and type of lipid has marked effects. Workers in the United States, Australia and New Zealand have reported on the effects of lipids, grazing and forage crops on lamb flavour. These will be considered in Chapters 1 and 12 since they form the main topic of this thesis. Paul et al (1964) did not always identify such differences.

Stress

Preslaughter stress depletes muscle glycogen reserves. Thus the extent of postmortem glycolysis will be reduced and the pH of the ripened meat will be higher than normal. Ovine meat from well fed unstressed animals is reported by Ford and Park (1980) to have a normal range of 5.4 to 5.8. The influence of ultimate pH on water holding capacity and hence juiciness and on mechanical properties was reported by Bouton et al (1972). Ford et al (1972 unpublished) reported on meat samples from 108 lambs of varying breed and sex. Meats were cooked/

cooked as stewed minces and served hot. Meat flavour intensity and hedonic ratings of flavour acceptability both showed highly significant negative correlation with ultimate pH. In these laboratories high scores in hedonic ratings are awarded to superior samples in contrast to the practice at QMC described in Chapter 3. Aroma intensity showed a smaller but highly significant positive correlation with ultimate pH. Thus intensity of lamb flavour decreases with increasing ultimate pH with a concurrent increase in non-meat character. These findings are of interest in that aroma is the most significant component of flavour.

Unpublished work by Murray (1976), also of CSIRO Meat Research Laboratories, Queensland, indicates that gross changes occur in the qualitative and quantitative composition of the flavour volatile fraction. Meaty characteristics of cooked lamb decrease with rise in ultimate pH thus producing different flavour character notes.

It will be of interest in the future to determine if pH values of gigot and loin samples used in the present trials and which showed such variation had any effects on hedonic rating scales, flavour preferences or the ability to detect differences in flavour.

The complexity of assessing lamb flavour is great. The chemical basis of meat flavour has not yet been established and its study presents a formidable undertaking. Since preslaughter factors, carcass ripening and cooking process all exert so much influence on flavour, it is perhaps not surprising that the reports from different laboratories on lamb flavour show such variation. However highly trained, a member of a taste panel cannot be considered to be merely an instrument. These factors combined with what appears to be variable eating quality of standard lamb samples, pose many problems in the assessment of lamb flavour.

CHAPTER 3

Summary

The rationale of sensory appraisal tests used to assess the effects of feeding regime on lamb flavour and the reasons for their selection are explained. Statistical techniques used in the analysis of results of sensory tests are indicated and their basis explained. Some of the points highlighted will be considered in greater detail in Chapter 12. Whilst much of the data was analysed with computer support a change from Systemshare to an ICL computer in 1979 has meant that later studies were carried out by visual inspection of printouts.

1. Sensory Appraisal Techniques and Statistical Methods Used in the Present Study

It was in 1949 that Boggs and Hanson made a critical appraisal of the analysis of foods by sensory difference tests. Whilst some of their references are now over 50 years old, their review was comprehensive and thorough and provides an interesting insight into the development of sensory appraisal techniques. Jellinek (1964) made a very thorough review of odour, taste and flavour evaluation with special emphasis on descriptive techniques. Amerine et al (1965) provided a text so comprehensive that it remains the standard reference almost 20 years later. In 1968, ASTM published a study correlating subjective and objective testing of odour and taste. In the United Kingdom British Standard BS 5098 (1975) and BS 5929 (1980) refer to a glossary of terms and methodology respectively. Meyer (1961), Griswold (1962), Paul and Palmer (1972) and Campbell et al (1980) refer to the sensory evaluation (organoleptic appraisal) of foods. Many successful symposia have been organised with this subject/

subject as the theme. More recently, statisticians rather than behavioural scientists have been concerned with the analysis of data from sensory appraisal tests. Information is thus readily available from a variety of sources.

Sensory appraisal tests are classified either as Difference/ Similarity and Descriptive. There is some overlap between the two categories.

Tests used in the present study and the basis for their selection are discussed below.

Difference testing is a very sensitive method of analysis which requires precision in test design, administration of the test and in the analysis of results. Simple (True) difference testing requires only either 'there is a difference' or 'there is no difference' response. Lack of ambiguity increases objectivity. Peryam and Schwartz (1950), Lockhart (1951) and BS 5929 (1980) emphasise that establishing that differences exist between samples should precede studies relating to magnitude or preference. Gregson (1960) suggested that subjects may be more ready to express preference in borderline discriminations. Thus he considered that consistency in two sample directional difference tests (Paired Comparison Preference tests) may provide better evidence of discrimination. The experimenter controls the characteristics and administration of stimuli and selects the technique to be used. The control of stimuli is discussed in more detail in Chapter 4. Results are analysed on the basis of the null hypothesis, i.e. that samples do not differ. Subjects either detect or do not detect differences. By selecting restrictive levels of statistical significance Type I errors are minimised and by using acute, reliable judges, in the present trials well motivated and experienced tasters, Type II errors/

errors are minimised particularly when the number of tasters is high.

That a difference may arise by chance alone is influenced by stimulus difference, the number of observations and variability between both subjects and trials. However as stimulus difference increases, the probability that judges will detect differences increases. The increased percentage of judges detecting difference between samples may be greater in one trial than another. Thus increased percentage responses are associated with greater differences between stimuli. True difference tests are one-tailed in that there is only one correct answer, whereas two-sample preference tests are two-tailed in that either response is valid. With meats, where homogeneous samples are more difficult to achieve, the number of presentations should be greater than in more simple food systems such as fruit juices.

The two-sample test selected for the present trials was the Paired Comparison (Preference) Test. Preference judgement on the basis of flavour was required. The nature of possible differences arising as a result of feeding regime could not be predicted in advance. Hence the use of Paired Comparison (Difference) tests, where a specified product attribute is required, was not considered desirable. Comparative adjectives such as 'blander', 'stronger' (Gridgeman 1955; Gridgeman 1970) could have been interpreted differently by individual tasters and the terms selected might have been inappropriate to differences between samples. A less well documented version of the two sample test in which tasters are required merely to state if two samples are the same or different (Greenhaugh 1970) was not considered to be appropriate in that Type 1 errors would have been likely to arise.

Preference/

Preference is highly subjective in nature. Hence four replicated paired comparisons were made. Tasters were judged to be consistent if either sample was preferred on three or four occasions. It is often assumed that preference judgements are made on a rational basis and that subjects who show consistent preference should also be able to identify differences between samples, that is that they are discriminators. As in the present series of trials, it is not only discriminators but also non-discriminators who demonstrate consistent preference. This problem will be considered later in this chapter.

Since two-sample difference tests could not be used in these trials, it was necessary to use either a three-sample test or a two-in-five test. At the time of the trials, it was believed that the triangle test was the most commonly used and researched of the three difference tests. In the triangle test, three samples are presented, two of which are identical. Subjects are required to identify the odd sample. The probability of correct identification by chance alone is $33^{1}/3\%$. Duo-Trio tests may also be used. One stimulus, identified as the standard, is presented first, followed by two different stimuli one of which is the same as the reference (standard). The subject is required to indicate which of the two stimuli is the same as the reference. In this test the probability of correct identification by chance alone is 50%. Thus the reason for the greater use of the triange test is perhaps explained particularly since it was considered to be the most studied and criticised of all test designs (Amerine et al 1965). More recently, the use of the Duo-Trio test is being investigated in that more samples may be assessed at a single session without inducing sensory fatigue. The power of the two tests is at present being assessed in the author's laboratory. There is a very good appraisal of the two tests by Harper in/

in Human Senses in Action (1972). Spencer (1979) compared the two tests in a single experiment. He indicated the advantages of the Duo-Trio and showed it to be more sensitive than the triangle test. The two-in-five test was considered to be too complicated for use in the conditions of the present trials.

Muller (1977) reporting on a survery of SCI Sensory Panel members and Research Institutes (208 and 10 : replies 53 and 7 respectively) indicated that assessors of meats do not usually have reference samples available. The majority of assessments were made by panels of one to ten people (87%) although it was indicated that most assessors are trained by a variety of methods. In both industry and Research Institutes the triangle test was more frequently used than the Duo-Trio. These tests differ from others reported in this survey in that they are used only to demonstrate if differences do or do not exist between samples whereas in most practical situations scoring overall quality or product attributes separately are the most commonly used techniques. Paired comparisons and ranking tests were also used. It therefore seemed that the use of triangle tests combined with paired comparisons was appropriate for the present study. In triangle tests positional bias has been recorded (Harrison and Elder 1950; Harries 1955; Frijters 1977; McBride and Laing 1979). The findings of different workers are not always in agreement. Presentation of the odd sample can occur in six different arrangements. A learning effect may occur when the odd sample remains constant in a series of trials. Thus to avoid positional bias, six randomised presentations is the procedure of choice. This procedure, whilst acceptable to the statistician, is impracticable in testing situations such as those described in Chapters 6-9 of this thesis.

In/

In the design of the trials described in Chapter 6, it was considered likely that the flavour of samples from rape fed lambs would differ from grass fed samples. Thus for triangle tests (aroma) only presentations AAB, ABA and BAA were used with B being the rape fed sample and A the grass fed (control) sample. Three batches of raw and three of cooked samples were presented. To minimise positional bias and possible learning effects, raw samples were assessed in trials 2, 3 and 5 and cooked samples in trials 1, 4 and 6. Even this series of presentations involved sniffing the contents of 18 stoppered tubes. Additional trials could have further increased the possibility of sensory fatigue. As will be indicated in Chapter 12 the procedures proved to be distasteful, difficult and destructive of tasters' confidence in their discriminatory ability.

In tasting meat samples, four paired comparisons were combined with either two or four triangle tests. Randomised arrangements of AAB or BBA were used in the triangle presentations. Thus either 14 or 20 samples were presented for assessment. Whilst retasting was allowed, the possibility of confusion was pointed out to participants who were encouraged to use palate cleansers.

Frijters (1979) indicates a problem which could arise in triangle tests where A and B samples are very similar in flavour. There will be an area of overlap in the normal distributions of internal sensory responses associated with stimuli A and B. Hence it could be difficult to identify the odd sample in the AAB or BBA presentation if either of the 'identical' samples (stimuli) overlapped with the odd sample. This is particularly a problem with such a nonhomogeneous food as meat (Land, personal communication) where the preparation of identical samples/

samples is so difficult. The difficulty of preparing identical samples was noted by the experimenter who has found preparation relatively easy with roasted pork but extremely difficult with roasted lamb (particularly gigots) with roasted beef in an intermediate position. This problem is possibly highlighted by the flavour profile analyses on the samples from rape and grass fed lambs (Appendix Table 8.1). Thus it becomes possible to suggest reasons for the violation of the assumption of the triangle test, i.e. the correct identification of the odd sample may be achieved by chance alone in $33\frac{1}{3}\%$ of presentations which was demonstrated in some of the trials. It could be concluded that differences between samples, if any, were marginal.

It had always been the intention to link the paired comparison and triangle test results for each subject. In this series of trials, consistent preference was not always associated with good performance in triangle tests (>50% correct identifications of the odd sample). Hence the view that discriminators (defined as above) were more likely to demonstrate consistent preference for either of the two samples was challenged. Greenhalgh (1970) investigating the extent to which triangle tests suppress discriminatory ability in tasters, noted that discriminators and non-discriminators were approximately equally consistent in their preferences. Despite modifications in his presentation techniques, it was not possible to modify findings. These results were obtained using medicines of different formulation. Perhaps his lower confidence in results of triangle tests cannot be applied to complex food systems such as roasted lamb. Using potato and corn crisps, Woodward and Schucany (1977) concluded that the use of non-discriminators' preferences tended to cause convergence in the preference/

preference for a particular sample. However, as already stated, the triangle test, despite possibly artificial constraints, was reported by Muller (ibid) to be the most commonly used difference test at the time of these trials. It had proved satisfactory in identifying differences in the experiments of 1975 and 1976 (Related experimental studies).

In addition to subjecting tasters to the stress of such experimental procedures, some workers also request them to indicate the magnitude of the differences between the odd and identical sample in the triangle test. In 1975 and 1976, in a series of 30 trials, testers assessed differences in the aroma of subcutaneous fats from pigs on a de-oiled herring silage regime from matched controls, by the use of triangle tests. They were requested to rate differences between odd and identical samples as 'slight','moderate' or 'great'. Results of these trials are given in Table 3.1. It should be noted that in none of the trials was the basic assumption of the triangle test violated and that statistically significant differences were demonstrated between samples in many of them. Responses are studied in more detail in Table 3.2 where the totals of correct and incorrect responses are indicated together with the related percentages of differences described as 'slight' and 'great' respectively.

Since the numbers of correct/incorrect responses varied, it was necessary to use these percentages for purposes of comparison. As might have been expected, more of those who identified the odd sample incorrectly described differences as slight. There was however a wide range of responses as reflected in the Standard Deviations. For this reason it is probably wise to consider that there was little difference between the two groups where difference between samples

was/

Table 3.1 Performance in Triangle Tests and Magnitude of Difference Estimates

Test No.	Correct (C)			Incorrect (I)		
	S	M	G	S	M	G
1	18	6	8	18	21*	6
2	16	10	4	22	21	5
3	16	12	9	22	12	3
4	16	5	7	20	13	4
5	19	16	5	18	14	3
6	14	20*	5	22	13	5
7	14	8	7	26	20	2
8	15	10	4	17	19*	8
9	12	11	8	22	11	3
10	24	21	8	14	8	0
11	12	11	7	23	16	3
12	21	7	3	7	9*	6
13	15	10	0	27	14	2
14	13	6	3	29	19	8
15	17	21*	16	10	9	3
16	7	1	1	18	9	1
17	12	6	3	8	10*	2
18	7	6	0	14	9	1
19	7	5	2	15	5	3
20	5	10*	17	3	3	3
21/						

Differences described as:

S = slight

M = moderate

G = great

Table 3.1 (Contd)

Test No.	Correct (C)			Incorrect (I)		
	S	M	G	S	M	G
21	11	11	2	10	5	0
22	4	8*	7	11	6	3
23	8	5	2	19	2	0
24	12	8	1	8	3	1
25	9	3	1	9	6	5
26	8	3	0	18	2	2
27	6	13*	6	6	9*	1
28	3	18*	5	6	3	1
29	9	14*	5	3	6*	1
30	6	7*	2	15	9	0

*Results where moderate assessments exceed slight.

Differences described as:

S = slight

M = moderate

G = great

Table 3.2 Comparing Estimates of Difference between Samples: Correct and Incorrect Judgements in Triangle Tests

Trial	Correct			Incorrect			Total in Trial
	N	%S	%G	N	%S	%G	
1	32	56	25	45	40	13	77
2	30	53	13	47	47	11	77
3	37	43	24	37	59	8	74
4	28	57	25	37	54	11	65
5	40	48	13	35	51	9	75
6	39	36	13	40	55	13	79
7	29	48	24	48	54	4	77
8	29	52	14	44	39	18	73
9	31	39	26	36	61	8	67
10	53	45	15	22	64	0	75
11	30	41	23	42	55	7	72
12	31	68	10	22	32	27	53
13	25	60	0	43	63	5	68
14	22	59	14	56	52	14	78
15	54	31	30	22	45	14	76
16	9	78	11	28	64	4	37
17	21	57	14	20	40	10	41
18	13	54	0	24	58	4	37
19/							

Abbreviations as in Table 3.1.

Table 3.2 (Contd)

Trial	Correct			Incorrect			Total in Trial
	N	%S	%G	N	%S	%G	
19	14	50	14	23	65	13	37
20	32	16	53	9	33	33	41
21	24	46	8	15	67	0	39
22	19	21	37	20	55	15	39
23	15	53	13	21	90	0	36
24	21	57	5	12	67	8	33
25	13	69	8	20	45	25	33
26	11	73	0	22	82	9	33
27	25	24	24	16	38	6	41
28	26	12	19	10	60	10	36
29	28	32	18	10	30	10	38
30	15	40	13	24	63	0	39

where N =total correct/incorrect judgements, and %S and %G = % rated as slight and great respectively

was described as great. Comparisons of the two groups are shown in Table 3.3.

Table 3.3 Comparing Estimates of Differences between Samples: Correct and Incorrect Judgements in Triangle Tests

	Slight		Great	
	$\bar{x}$	SD	$\bar{x}$	SD
Correct	45.6	18.1	16.9	11.2
Incorrect	54.6	14.7	10.3	7.8

Following the study of this data, it was decided that the distraction of requesting tasters to categorise differences between the odd and identical samples was undesirable particularly when even results demonstrating detectable differences were so inconclusive.

Thus in the majority of trials, paired comparisons (preference) and triangle tests were used. It was always recognised that if detectable differences were demonstrated by triangle tests that further tests would be required to determine their nature and magnitude. Their nature might also be of importance in relation to tasters' preferences.

2. In analysing data from the paired comparisons and triangle tests, the theoretical basis for statistical tests of significance was based on their binomial distribution (Pridmore 1979). Classical statistics are based on the assumption of normality for populations sampled. Non-parametric methods avoid this assumption and may be of greater efficiency when results are not normally distributed (Hollander and Wolfe 1973). The importance of such techniques, whilst recognised by the statistician, is not widely recognised by scientists other than behavioural scientists. Many of the sensory tests used by the food scientist are equally relevant and appropriate to the behavioural scientist. Hence increasing co-operation/

co-operation between statisticians and food scientists in the design of experiments and appropriate techniques to analyse data is encouraging. Major developments in non parametric statistics have occurred during the last 45 years.

A specimen recording form for the tests is included in the Appendix (Table 6.3). As indicated in Chapter 6, whilst the same form was used to avoid punching errors when in results of the trials, the actual presentations to tasters varied from session to session and the aroma trials were carried out at random. Table 6.2 indicates the experimental design. Probabilities were calculated in triangle tests on the basis that a trial is one set of either two, three or four triangle tests. Each judgement, having a one in three chance of being correct, cannot however be regarded as an independent event since each subject carried out the procedure using the same samples on more than one occasion. Replications allowed comparisons of assessors' performance to be made (Pridmore 1979 ibid).

For both paired comparisons and triangle tests, on the null hypothesis, results are binomially distributed. Thus the observed and expected frequencies may be compared. Values of χ^2 with df (K - 1) are calculated. Statistically significant flavour preferences and differences in flavour and aroma could thus be determined.

The Kolmogorov-Smirnov one sample test was used to study differences in the observed and expected frequencies. This test is concerned with the degree of agreement between the sample scores and theoretical frequencies, that is, it is a test of goodness of fit. It determines if sample scores are drawn from a population having the same theoretical distribution. This test should be distinguished from the Kolmogorov-Smirnov two sample test which is concerned with the agreement between the/

the distribution of two sets of sample values rather than comparison with a theoretical distribution. Literature studies suggests that the two-sample test is more frequently studied and used. Biometrika Tables for Statisticians, Volume 2, 1972 reprinted with corrections (1976), provides only tables related to the two sample test. Pratt and Gibbons (1981) indicate that the one-sample test can be used either to test the null hypothesis using cumulative distribution functions (c.d.f) or to determine confidence regions. It differs from Pearsons's χ^2 goodness of fit test in that it determines cumulative rather than cell frequencies. The section in their text is brief in comparison to the information on the two-sample test.

Cumulative frequency distributions of the theoretical and observed values are compared. The point of greatest divergence, known as the maximum deviation (D), is determined. Critical values of D for differing sizes of sample (N) are indicated in Table E (Siegel 1956).

To carry out the test, a table is constructed indicating the following:

F = number of subjects in a particular category

Fo(X) = theoretical cumulative distribution of scores

Sn(X) = cumulative distribution of observed scores

Fo(X) - Sn(X) for each category is then calculated. The value of D is thus identified. If this value of D is the same as or exceeds the critical value of D in Table E the null hypothesis may be rejected.

The Kolmogorov-Smirnov one-sample test treats individual observations separately. In the χ^2 test information may be lost by grouping categories with consequent loss of sensitivity. Thus in certain conditions the Kolmogorov-Smirnov test may be the more powerful test./

test. Should the null hypothesis be rejected, in addition to examining the data, further tests are required to determine the nature of the observed frequencies deviation from the theoretical distribution.

In addition to the use of the expanded tables for estimating significance in paired comparisons and triangle tests (Roessler et al 1978), the actual probabilities of results being obtained by chance alone were calculated by the use of the z statistic. For paired comparisons, where N >5,

$$z = \frac{(X1 - 0.5) - m}{\sigma}$$

where $m = {}^{n}/_{p}$ and $\sigma\sqrt{npq} = \sqrt{m/2}$

and $p = q = \frac{1}{2}$ X1 = correct responses Formula 3.1

Since paired comparison (preference) tests are two-tailed, the probabilities obtained by the use of Appendix Table A (Amerine et al 1965) must be doubled to obtain the actual probability.

Triangle tests are one-tailed and

$$z = \frac{(X1 - 0.5) - m}{\sigma}$$

where $m = np = n/3$ and $\sigma = \sqrt{npq} = \sqrt{nx^{1}/3 \times {}^{2}/3}$

$$= \sqrt{2n/9} = \sqrt{2n}/3$$

where p = odd sample and q = two identical samples

Formula 3.2

In Chapter 6, the hypothesis that it would be easier to detect differences between raw grass and rape fed samples than in cooked was tested by the Wilcoxon Matched-Pairs Signed-Ranks Test.

This is a more powerful test than the Sign test in that it utilises not only the direction of differences between pairs but also takes account of the magnitude of these differences. Greater weight is thus/

thus given to pairs which show a large difference between the two conditions. The number of correct identifications (0, 1, 2 or 3) for raw and cooked samples was recorded for each tester. The difference between scores was recorded in this way.

Subject	Raw	Cooked	Difference
A	3	2	-1
B	2	2	0
C	1	3	+2

Subject B, where the scores were equal was dropped from the trial. The scores of subjects where there was difference in performance were then ranked. T is the smaller number of like-signed ranks, i.e. either the sum of positive or negative ranks whichever is smaller. N is the number of pairs of differing sign. If the observed value of T is equal to or less than the value quoted in Appendix Table J (Haber and Runyon 1977) the null hypothesis can be rejected. Since an advance prediction of the result was made the test was one-tailed. For large samples where N >25, the z statistic is calculated. This procedure was not required in the present trials.

For small samples, the efficiency of the Wilcoxon Matched-Pairs Signed-Ranks is approximately 95% in comparison with the same data analysed by the parametric t test (Wilcoxon 1945).

These statistical tests were used in the majority of the present trials. The use of Friedman's Two-Way Analysis of Variance, Spearman's Rank Correlation Coefficient and the Sign Test in the experiments described in Chapter 5 will be considered later in this chapter as will the Median Test used for the analysis of data in Chapter 10.

3. In trials to determine the optimal internal temperature for roasted/

roasted lamb (Chapter 5), hedonic rating scales were used. In hedonic scaling affective responses of like and dislike are measured. Scales differ in the type and number of aids or cues and fineness of discrimination expected, but all require assignment in ordered categories along a continuum. After this scores are allocated. They cover a range of affective responses which are extremely pleasant through indifference to extremely unpleasant.

Interpretation of these responses varies. A numerical value of zero may be assigned to the indifferent category with positive (like) integers above and negative (dislike) integers below. Whilst the unsophisticated could be confused by the break in the continuum of the scale, Amerine et al (1965) considered that such a scale could be used effectively with judges who have been trained to understand the terms which are used. Rhodes (personal communication) does not accept the indifference of judges. On this type of scale he considers the judgement allocated zero to be invalid (see also Harries et al 1963).

In other laboratories, only positive integers are used. Following the pattern of conventional scoring techniques, the highest scores are allocated to the most pleasurable responses. On the other hand, some workers consider that subjects are required to be more attentive if pleasurable responses are allocated the lowest scores. In the author's laboratory the allocation of low scores to pleasurable, i.e. most preferred responses has been the practice for many years. The technique has been standardised for all experimental work so that no confusion arises. In analysing data, there is no difference between the two practices provided that it is quite clear how scores have been allocated. Examples of the hedonic rating scales and response sheets are included in the Appendix (Tables 3.1 and 3.2).

Choice/

Choice of words is extremely important. Ambiguity must be avoided. Clarity appropriate to educational background is essential. Ellis (1968) describes the use of several facial hedonic rating scales for use with young children and for subjects with poor verbal comprehension. The selected words must convey the idea of the successive order of the categories and make clear the meaning of the response continuum.

Jones et al (1955) assessed the time taken, reliability, accuracy in duplicated tests and the amount of information acquired about foods using 900 enlisted soldiers. Nine scales were investigated. None was demonstrated to be superior. It was however noted that nine interval scales tended to be more sensitive to differences between foods than shorter scales and that an equal number of positive and negative categories is not an essential feature.

The historical development of hedonic rating scales for foods was traced by Peryam and Pilgrim (1957). A seven point hedonic rating scale was first used by the Quartermaster Food and Container Institute in 1947. Examples are illustrated in Amerine et al (1965). For laboratory use a vertical scale was selected as the format likely to suggest the idea of a continuum with equidistant points. A horizontal format was used with subjects who had had no previous experience in food testing. Since its use was apparently successful it can be assumed that it was clear and unambiguous. The method was used to detect small differences in the degree of liking for similar foods. Subjects were urged to respond on the basis of immediate feeling. Contrast effects posed problems with certain foods such as coffee but not with fresh milk or fruit juices. Contrast effects with/

with a 'good' sample following a 'poor' are probably inevitable when samples are tested simultaneously as in the present trials. Thus in testing roasted meats it was recognised that unless samples show considerable variation only small differences in degree of liking are to be expected. Cover (1959) considered that quantitative scales can be used for assessing tenderness and juiciness of meat samples but that responses will differ between judges who prefer rare and who prefer well done meats.

The psychological error of central tendency can explain the reluctance of subjects to use the extremes of any scale with resulting contradiction of the range of categories. This is particularly true of inexperienced subjects. Thus a nine point scale becomes effectively a seven point scale with seven and five point scales becoming similarly less sensitive. As a result of the author's work with students, a five point scale was considered too insensitive and a nine point scale too confusing for the less inexperienced. A seven point scale - as used by the Quartermaster Food and Container Institute (ibid) - has been used at QMC for many years and was selected for the present study. Lower numbers were allocated to favourable affective responses. The response sheets and the hedonic rating scale are included in the Appendix (Tables 3.1 and 3.2). Miller (1956) has indicated that stimuli can be assigned by most subjects to five and seven categories. Howgate and Smith (1981), reporting on work at Torry Research Station, also used a seven point scale.

It is recognised that responses are highly subjective and individual. Many factors contribute to variation in response. Where a score of three is awarded, even an experienced assessor could equally have judged the score to be two or four. Hence differences between samples/

samples must be relatively great and participants exceed ten for meaningful results to be achieved. Even when these criteria are met the highly subjective nature of responses can cause problems. This difficulty was considered by Piggott and Land (1981).

Major advantages of hedonic rating scales are simplicity, suitability for a wide range of participants since no previous experience is required. Results are an indication of general levels of preference. In addition, ratings require less time than paired comparisons or triangle tests, the procedure is of interest to participants and large numbers of stimuli can be presented. Whilst experience is an advantage in that judgements can be anchored it is not an essential prerequisite.

Disadvantages, in addition to variability in responses already described, include the need for relatively large numbers of participants to provide precision and that despite careful control of variables responses may change. Shephard (1955) suggests that if scores and descriptive terms are used simultaneously some participants may rate primarily by scores thus using a linear scale. Amerine et al (1965) observe that whilst judges initially anchor their evaluations by the use of descriptive terms, they may subsequently proceed to score numerically thus utilising a linear scale. Shephard (1954) also considered this point. These discussions acknowledge that more evidence of the applicability of parametric statistical techniques such as analysis of variance to scores assigned by the use of descriptive words is required. If parametric statistics such as means and standard deviation are used, it is implied that descriptive terms used represent equal sensory intervals and that scales are linear. Amerine et al (1965) acknowledge disagreement among investigations as to whether or not equally spaced grade numbers/

numbers of arbitrarily ordered descriptive terms imply equal intervals of quality. More detailed discussion of these points is deferred until Chapter 12.

4. In the present study data derived from the use of hedonic rating scales has been analysed by nonparametric techniques. Hedonic rating scales assessed subjects' attitude to lamb flavour. Whilst the essence of the method is simplicity, subjective assessments rather than precise measurements are achieved.

Thus hedonic ratings form a classificatory ordinal scale with the property of rank ordering. It is thus possible to determine whether lamb flavour possesses more or less of a particular characteristic but there can be no indication of the magnitude of differences or of their nature. There is however, a relationship between the categories in that samples are classified on the basis of a 'more preferred' or 'less preferred' response. This relationship applies to all pairs of classes so that a complete rank ordering is possible. Hedonic rating scales should therefore be considered to be ordinal.

In analysing data, scores are assigned to each category. These scores may appear to be more precise than ranks but they do not meet the criteria for higher levels of measurement. Transformations which do not change the order of ranks are therefore admissable. No loss of information is thus involved.

The statistic most appropriate to describe the central tendency of scores in ordinal scales is the median. The median is not affected by changes of any scores above or below as long as total scores in both categories remain the same. The use of the median may compensate for alteration in responses during trials despite control of variables reported earlier in this section. When only rank order of scores is known/

known, the use of arithmetic means and standard deviations are suspect in that successive intervals between categories are not equal. The properties of an ordinal scale do not allow such arithmetical procedures and hypotheses based on such measurements must be viewed with caution particularly if response patterns in a trial do not approximate to the normal distribution.

In the Friedman two-way analysis of variance by ranks, scores are first cast in a two-way table with n rows - where n is the number of tasters - and k columns - where k represents the number of samples. In Chapter 5, the four columns represented lamb samples cooked to internal temperatures of $70^{o}C$, $75^{o}C$, $80^{o}C$ and $85^{o}C$. Each row represents a given taster's performance. An example of such a two-way table is given in Appendix Table 5.4. Table 5.5 of the Appendix shows the same information when the calculated values in sign tests were checked by the use of a newer program. This program calculates median values and column totals. In the earlier program means and standard deviations were calculated purely for interest. Their limitations were recognised. Medians were subsequently determined by inspection rather than with computer support (Chapter 5).

In the first stage of analysing data, scores in each row, that is for each taster, are ranked. Allowance is made for ties. As well as being appropriate for ordinal data, ranking assists in overcoming the problem of less experienced tasters who consistently award either higher or lower scores than their contemporaries. Ranks in each row range from one to four. If the null hypothesis applies, that is that all samples irrespective of internal temperature elicit the same response, the distribution of ranks in each column will be such that ranks one, two, three and four appear with an approximately equal frequency./

frequency. For an individual taster, it is a matter of chance as to which samples are awarded the highest or lowest scores. Hence rank totals (CT in Tables 5.4 and 5.5 of the Appendix) for each column will be approximately equal implying that, in the opinion of the tasters, samples of lamb did not differ. However, if the null hypothesis is rejected, samples differ and rank totals will thus vary from one column to another. This is analagous to the situation in parametric statistics where it could be stated that the mean values of the ranks in each column would not be equal.

The Friedman test determines whether the rank totals (Rj) differ significantly. To carry out the test, the value of the statistic χr^2 is calculated. When the number of rows and/or columns is at least nine and three respectively, Friedman (1937) demonstrated that χr^2 is distributed approximately as χ^2 with df = K - 1. The value of χr^2 is calculated as follows:

$$\chi r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (Rj)^2 - 3N(k+1) \qquad \text{Formula 3.3}$$

where N = number of rows, k = number of columns,

Rj = sum of ranks in j.th column

$\sum_{j=1}^{k}$ requires that the squares of the sums of ranks over k conditions be summed

In the present experiments where k = 4 and N always exceeded 9, χr^2 can be considered to be distributed as χ^2 with df(k-1). If the calculated value of χr^2 equals of exceeds these figures the null hypothesis is rejected. Thus a statistically significant difference is demonstrated between the sum of ranks (column totals) in the four columns. Alternatively, it may be considered that there is a statistically significant difference between the mean ranks Rj/N of the/

the four columns.

In Friedman's paper (1937), the χr^2 was compared with the parametric F test. In a series of 56 tests, 45 probability levels were essentially the same. For the 56 sets of data the null hypothesis would have been rejected on 26 and 24 occasions ($p<0.05$) in the χr^2 and F tests respectively. The two tests would therefore appear comparable in power and the χr^2 test has the advantage that it is not assumed that there is an equal interval between categories (Friedman (1940).

It is accepted that parametric techniques are still widely used by many experimenters in the analysis and interpretation of hedonic scale data. The validity of this practice is being challenged. Many who use parametric statistics accept their limitations but continue to use them. There is further discussion of these points in Chapter 12.

If a statistically significant difference between column totals is demonstrated, it is usually clear where the differences are to be found. To confirm the observation or, where required, to identify the difference, the Sign test may be used. Only plus or minus signs are used. Quantitative differences are ignored. The test applies to two related samples. The aim is to establish if differences exist between them.

From Tables 5.4 and 5.5 of the Appendix, there are four columns corresponding to varying internal temperatures. If these columns are designated A, B, C and D, tasters scores on two tests are successively compared; that is AB, AC, AD, BC, BD and CD. For example, if the score for B exceeds the score for A, a plus is recorded. Conversely, if the score for B is lower, a minus is recorded. Ties are discarded from the trial. The number of remaining pairs, irrespective of sign is/

is designated N. Table M (Haber and Runyon 1977) indicates the one-tailed probability associated with a value as small as the observed value of x, where x is the number of fewer signs. In these trials, the region of rejection is two-tailed since the direction of differences was not predicted in advance. Hence the probability values in Table M were doubled.

Results of the Sign tests are given in Table 6.6. As indicated in Chapter 6, they assisted in the selection of optimal internal temperature. Whilst the Sign test is less powerful than the Wilcoxon Matched-Pairs Signed-Ranks test described earlier in this chapter, it provides a simple and rapid method of identifying differences. When Median tests were carried out there was little difference from the results which were obtained using the Sign test.

The Spearman Rank Correlation Coefficient, r_s, sometimes known as Rho, is used to compare two ordinally scaled variables. Thus Spearman Rho values were calculated for duplicated hedonic ratings to assess tasters' consistency in judgements and to compare performance when samples were tasted hot and cold. Results are presented in Tables 6.2 and 6.3 of the Appendix.

The formula used in the calculation was:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}} \qquad \text{Formula 3.4}$$

where $\sum x^2 = \frac{N^3 - N}{12} - \sum T_x$

$$\sum y^2 = \frac{N^3 - N}{12} - \sum T_y$$

Values of x and y are the ranked scores for the first and second tests respectively. The correction factor, T makes allowance for ties.

$T = \frac{t3 - t}{12}$ where t = number of scores tied at a given rank

The/

The $\sum T$ is the sum of the various values of T for all the groups of tied observations. In Formula 3.4, $\sum d^2$ represents the sum of the squares of the differences between x and y scores for each taster. Differences are squared so that negative and positive values of d do not cancel each other out. It is thus possible to determine the magnitude of the disparity between scores.

The statistical significance of the test for between four and thirty tasters is determined by the use of Table G (Haber and Runyon 1977). If the observed value exceeds the tabled value the result is of statistical significance. The test was one-tailed in that it was predicted that there would be a positive correlation between the results of the two tests.

To summarise, the ranked scores on the x and y variables were tabulated. The value of d, for each taster was obtained by subtracting the y rank from the x rank. Values of d_i^2 were summed to give $\sum d_i^2$.

The power efficiency of the test was judged to be 91% in comparison with the correlation coefficient r for linearly related data (Hotelling and Pabst 1936). Its use is still recommended in more modern tests such as Haber and Runyon (1977).

As indicated earlier in this section in nonparametric statistics, the median is the appropriate measure of central tendency for ordinal data. It indicates if groups differ in central tendency. Medians were extensively used in Chapter 5 to determine optimal internal temperature for gigots and loins. They were of value in confirming results of Sign tests and 'degree of liking' in selecting internal temperature of $75^{o}C$.

5. One sample Median tests were extensively used in analysing data in Chapter 10. They are considered in context. The combined median for/

EDINBURGH UNIVERSITY LIBRARY

for the two samples was determined. Values in each group were dichotomised as those values at or below the combined median or those above the combined median. The null hypothesis was tested using Formula 3.5 since samples sizes were large.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(Oij - Eij)^2}{Eij} \qquad \text{Formula 3.5}$$

where Oij = observed number categorised in the ith row of the jth column and Eij indicates the expected frequency. The sum over all rows (r) and columns (k) for each cell gives the value of χ^2 with df $(r - 1)(K - 1)$, that is one.

Further details of these non-parametric statistical procedures is given in Siegel (1956), Bock and Jones (1968), Hollander and Woolfe (1973), Kendal and Stuart (1979) and Pratt and Gibbon (1981). Blakesley (1977) is also of assistance. Stevens (1970 and 1974) also presents information.

6. In Chapter 8, it is indicated that Flavour Profile Analysis was used to confirm that there was no difference between roasted gigots and loins from grass and rape fed samples. Details of the technique are provided by Amerine et al (1965) and in papers by workers at the Arthur D Little Research Co., Cambridge, Masachusetts. Essential features are summarised.

Unlike difference testing, the profile method is not concerned with single flavour judgements which can be analysed statistically. The method concentrates on entire flavour and individual attributes of flavour in relation to each other. All flavour components are thus considered in perspective. Acceptance is not measured. It is considered by the experienced to be an almost completely objective sensory technique.

Four/

Four to six highly trained judges are used. Stringent selection procedures and a six to twelve month training period are required. Trained panel leaders organise, conduct and direct panel sessions and interpret results: they are the link between the panel and the user of panel findings. Panel discipline is strict.

In formal panel sessions, each panel member independently examines samples and records findings. This is followed by a group discussion led by the Panel Leader who subsequently tabulates and summarises the discussion. Disagreements are resolved by resubmitting samples. Flavour profile analysis includes perceptible aroma, taste, flavour and feeling factors described as character notes. Degree of intensity and order in which factors are perceived is recorded. After-taste and amplitude, that is overall impression, of aroma and flavour are graded. Each character note in aroma, flavour-by-mouth and aftertaste is identified and described. Findings may be summarised diagramatically. A semicircle denotes threshold concentrations with radiating lines for each individual character note in order of appearance. The length of the line represents intensity (Amerine et al 1965, p.382).

A major advantage claimed for flavour profile analysis is reproducibility. Caul (1957) reported identical profiles from the same soup tested initially and again after a year. Such claims are however difficult to substantiate in the absence of statistical analyses.

The latter is a major disadvantage, which combined with cost of training panel members, lack of precision of the intensity scale, inability to quantify individual sensitivity to specific odours, tastes or flavours and inherent errors of open discussion techniques, has led to modification of the techniques for testing specific products. In particular, it should be recognised that a group judgement is not the same as a group of/

of judgements. The leader or an individual panel member may influence the result. There may be unconscious signalling. Where flavour components are clearly defined, as for beverages, personal judgements are probably preferable but where flavour notes are as variable as with meats, open discussion is quicker and more flexible provided its possible limitations are appreciated.

The author does not agree with Tilgner's assertion that a dilution profile technique (1962) is the most satisfactory method to characterise sensory properties of meats. They should preferably be assessed as they would normally be eaten.

Table 8.1 of the Appendix is not a typical flavour profile analysis. Only two highly trained panel members were used. Both have acted as panel leaders. Panel discipline and testing conditions were as specified. However, they were able to confirm that it was not a matter of surprise that meats from two different feeding regimes demonstrated no detectable flavour differences.

7. At the time of the trials to determine optimal internal temperature (Chapter 5) QMC had a telephone link with a Honeywell Series 60 computer at Systemshare, Edinburgh. Digital signals were converted to amplitude modulated signals using a modem. At the main computer, a second modem reversed the process. Data was stored in files ready for analysis using appropriate programs. Printouts of results were obtained from a desk-writer at the QMC terminal. Although the Data Processing Centre was established only in 1976, despite some increase in facilities, it became obvious in 1978 that resources were inadequate.

In July 1979 an ICL 2903/40 was installed. To retain data for the lamb trials and tasters' performance in fundamental sensory tests/

tests paper tape was used. It is a matter of regret that the two systems were incompatible and that files could not thus be transferred to the ICL computer.

This has meant that monitoring of testers' performance over a series of trials and correlation between other sensory parameters must be carried out by inspection of printouts rather than with computer support as had been hoped. Further study of data with computer support is possible for experiments carried out during and since 1979 using programs adapted from those devised for earlier test procedures.

In conclusion, the reader is referred particularly to Amerine et al (1965) and BS 5929 (1980) for more detailed information on sensory appraisal tests and to statistical texts quoted earlier in the chapter for additional background information on the procedures described.

CHAPTER 4

Summary

The selection and preparation of lamb joints for cooking is described. Reasons for the choice and details of the cooking method are indicated. Presentation of samples to tasters to ensure uniformity, criteria for selecting tasters and tasting procedures are discussed. Throughout the importance of standardisation is emphasised. Some experiments carried out to determine tasters' sensory acuity are described.

Establishing a Methodology

Following the review of factors influencing meat flavour in Chapter 2, it is evident that meats show intrinsic variation. The meat technologist is aware that no two samples are exactly alike but may not fully appreciate that scientific control of the cooking process is essential in studies of flavour. Similarly those who study the effects of the cooking process on meat flavour may ignore differences between meat samples. This dichotomy of interest is evident in the literature. In planning experimental studies both aspects must be considered. Causes of variation in the meat itself such as breed, age, grade, pre- and post slaughter treatments, cut and culinary preparation must be standardised. At the same time, cooking conditions, although difficult to control, must also be standardised. The author had had some experience in controlling experimental variables in the two sessions preceding these investigations. Hence it was considered necessary to devise a methodology, appropriate to the United Kingdom, to minimise these variables so that the effects of feeding regime on lamb flavour - if any - could be determined. The procedures which were followed are described.

1./

1. Breed of Sheep

Scottish Blackface wether lambs from the Edinburgh School of Agriculture farms were used in all the trials except those described in Chapter 8. As indicated in this chapter, these were Suffolk x Dorset/Finn females born in January 1977. They were housed throughout until slaughter in May. After weaning in March, they were fed on a whole barley/protein pellet mix with a small supplement of hay. Joints from these lambs were used to assess tasters' acuity and methodology of the experimental procedures so that it was appropriate that they should be of a different breed and sex from the other lambs used in the trials. Joints from these lambs were compared with joints from much older Blackface wethers to determine if detectable differences could be demonstrated between them.

The Blackface lambs used in the main trials were born in April/May 1977 or April/May 1978. They remained on the hill until August. At weaning, they were grazed on lowland grass fields until they were allocated to their feeding regimes. Details of finishing diets are given both in context and in the Appendix. As far as possible, comparable lambs were selected for the trials.

2. Preparation of Lamb Joints

Preslaughter conditions were arranged so that joints were prepared four days post-slaughter in the Carcass Evaluation Unit. Feeding regime and the cooking process could have differing effects on lean and fatty tissues. Hence joints from gigot and loin areas were selected as examples of commonly used lean and fatty cuts. The proximate composition of the two cuts (cooked) is:

	Protein (N x 6.25)	Fat	Water
Gigot	29.4	8.1	61.8
Loin	27.8	12.3	58.9

Source: Paul and Southgate 4th ed. 1978 The Composition of Foods. McCance and Widdowson 4th ed.

Figures/

Figures are quoted for cooked lean only; in both cuts visible fat was removed before presentation to tasters.

Anatomic locations were standardised. The lower half of the carcass was split at the vertebral column. Gigots were first cut straight across in the patellar region. A second cut, parallel to the first incision, was made 40mm proximal to the first. For loins, Lumbar vertebrae 7 - 12 were used. Left and right sides were handled separately.

After preparation, joints were packed separately in aluminium foil containers and covered with foil/waxed cardboard laminate. Each sample was labelled to indicate weight (g), carcass number, anatomical location and date of preparation. Samples were then frozen until use. Storage temperature was $-25^{\circ}C$ ($+/-2^{\circ}C$). In later trials, the cost of the containers became so high that heavy duty labelled freezer bags were used. After labelling, the bags were enclosed in a tightly fitting labelled doubled layer of heavy duty aluminium foil. There was no evidence that this practice affected samples even after prolonged storage. There was no suggestion of fat rancidity. Weight change of thawed joints was always less than 8g compared with the original samples.

Many experimenters have used small, precisely shaped samples of muscle tissue enclosed in glass, metal or plastic heated in water, steam or oil to achieve better control of external temperature and heat transfer within samples. For studies concerned with detailed heat-induced physical and chemical changes these methods are entirely appropriate. However, considerable interpretation and modification of such experimental results is required before they can be applied to current cooking practices at the point of presentation to consumers or members/

members of a taste panel. This approach was considered inappropriate to a study where the aim was to assess consumers' responses to lamb flavour. Realistic cooking procedures for meat cuts readily available at retail levels was considered desirable.

The decision to use gigot and loin joints of standard anatomical location was reached on the basis of pilot studies in July 1977. In these studies, chump, gigot and loin chops from the left and right sides of a carcass had been cooked to internal temperatures of 70°C, 75°C, 80°C and 85°C. Total, drip and evaporative losses were calculated. Raw and cooked pH values were recorded. Techniques were used as described in Chapter 5.

A seven point hedonic rating scale was used to assess eating quality. Meats were tasted both cold and hot. Results were analysed with computer support. Files DA* were created for each batch of samples. Using Friedman's two-way analysis of variance, the value of Xr^2 was calculated. No statistically significant difference was established between the meats cooked to the four internal temperatures or between samples tasted hot and cold.

Since this was a feasibility study, results, although available, are not set out in detail here and were not used in calculating data described in Chapter 10.

Following these experiments it was agreed that larger portions of meat were needed to provide greater comparability and more reliable estimates of weight losses and pH changes arising as a result of the cooking process. In addition, a larger yield of meat would be required to enable more tasters to assess the meats and to allow the outer borders to be discarded to ensure uniformity of presentation. Thus the basis for the experiments described in Chapter 5 was established and it was decided that only gigots and loin samples would be used.

3./

3. Preparations for Cooking Joints

It had been intended to thaw samples removed from the freezer at 17.00 hours in a refrigerator cabinet overnight for cooking the following morning. This procedure proved unsatisfactory in that joints tended to remain frozen. It proved necessary to thaw them overnight at room temperature (18 - 20°C). They were placed on a trivet fitted over a disposable plate and covered with Saran (polyvinylidene chloride).

Surface moisture was removed the following day. Joints were reweighed. As indicated earlier, weight changes during frozen storage and thawing were always less than 8.0g. This was in contrast to beef and pork samples of comparable weights in other experiments. A sample of raw fat and lean of standard anatomical location was removed for the aroma tests. Similarly two standard 1.0g samples of lean were removed. These were pulverised and mixed with 5.0ml deionised water. After ten minutes, pH values were recorded using a calibrated pH meter. Separate scalpels, tiles and beakers were used for each sample. Joints were then reweighed. A sample of the recording forms which were used are included as Tables 4.1 and 4.2 of the Appendix.

4. Choice of Cooking Method

Pauline C Paul reviewed meat cookery comprehensively under the heading 'Influence of Heating Methods' in Cole and Lawrie (1975). The complexity of discussing and comparing findings between laboratories and of minimising experimental variables is noted. State Agricultural Stations and the U.S. Department of Agriculture established an informal group during the period 1920 -40. Findings were summarised in 1942. These recommendations were tabulated together with those of the American Home Economics Association (AHEA) Terminology Committee (Handbook of Food/

Food Preparation, 6th Edition, published in 1971). Roasting, grilling pot roasting, frying, braising, stewing and pressure cooking are considered.

The thirty intervening years had seen the development and use of temperature controlled grills, friers, roasters, slow cookers, microwave cookers, aluminium foil and transparent film wraps. Paul recommended that selection of heating methods should take account of previous research results which show the importance of controlling variables such as choice of cut of meat, containers, type of energy supply, heat-transfer medium, heating time and temperature, internal temperature end point and total time requirement. Meticulous attention to such details should increase the probability of achieving satisfactory research results. For example, the tightness of a foil wrapper in roasting determines whether it acts as an insulator or not. Tightness could be difficult to control. Differences in methods of heat transfer, size and shape of containers could cause obvious variation in results.

In the present study, it was not considered appropriate to use a microwave cooker. Meat cooked by dry methods differs in flavour from meat cooked by moist methods. The environment surrounding the meats during the cooking process differs: in the former it is a combination of melted fat and in the latter of simmering aqueous liquid. These differing environments combined with differences in cooking temperature produce varying flavouring compounds. In dry methods there is greater aldol formation and condensation, Maillard (carbonyl-amine) and caramelisation reactions increase. Decomposition products of fats polymerise. A study of the nature of these flavour differences is complex. Paul (1972) tabulated a wide variety of methods (60) which have been used in experimental/

experimental studies including immersion of samples in plastic bags in a waterbath. The reader is referred to the chapter on meat cookery in Paul and Palmer (1972) for the principles underlying each of the methods described.

In the United Kingdom, gigots and loins of lamb are traditionally cooked by dry methods, that is by roasting (baking), grilling or frying. Thus it was decided that roasting the joints would produce the most palatable samples whilst at the same time minimising experimental variables.

Larger cuts of meat can show considerable variation in internal temperature depending on the location of the sensor. This was much less of a problem in the present study where the joints were relatively small and the position of sensors was standardised and their position was checked to make sure they were not displaced during the cooking process. The temperature rise* when meat was removed from the oven in this series of trials was minimal. This is likely to have arisen from a combination of the cooking temperature selected (177°C) - where the external to internal temperature (75°C) gradient was low - and the initial weight of the joints (500 to 650g approximately).

By definition, a dry method of cooking implies that moisture may evaporate freely from all surfaces of meats. Thus joints were placed on a trivet and roasted uncovered.

Because of differences in flavour of meats cooked by moist methods, had these been used, it would have been difficult to transfer findings to meats cooked by more traditional United Kingdom methods. For instance, if meats are pounded to minimise textural differences, effects of the cooking process alter. If such pounded meats are subsequently stewed or converted to meat loaves to allow uniformity of/

*i.e. rise in internal temperature

of presentation in respect of appearance, texture and temperature, this is an unrealistic procedure for gigots and loins of lamb. If meats are braised in the absence of seasonings which could mask slight differences in flavour, this is also unrealistic and likely to produce poor motivation and unfavourable responses in tasters.

Control and monitoring are essential to research studies to minimise experimental variables. This is in contrast to consumer studies where variables are seldom controlled. Frying and grilling were therefore discarded as cooking techniques because these processes are difficult to monitor and to control. In addition, they are suitable only for small, thin cuts of meat where surface area is relatively large Flavouring compounds are concentrated on the surface. Appearance differs. These layers must therefore be discarded during the preparation of uniform samples for tasting. Wastage is thus relatively greater.

As a result, it was decided that joints to be compared should be roasted on the same shelf of a preheated oven. The Pilot Study (Related Studies) had already indicated that cross contamination between samples was unlikely to occur. Accurate control and monitoring of external and internal temperatures was therefore practicable.

5. Cooking the Joints

Weighed joints were placed on a trivet in a weighed standard shallow roasting tin. Each pair of samples to be compared was placed on the same shelf of an electric oven preheated to a temperature of 177°C (350°F). Whilst a temperature of 163°C is recommended in the United States and other research laboratories, as indicated in chapters which follow, a temperature of 177°C was considered to be more/

more representative of U.K. practice and would allow traditional time/temperature relationships to be maintained. It is recognised that even within the U.K. practices vary but it was necessary for experimental procedures to be standardised.

It should be noted that Mottram (1981) considered that dry methods of cookery were preferable in the preparation of meats for sensory and instrumental assessments. Roasting is preferable to grilling. Cooking to a fixed internal temperature is superior to cooking for standard times.

Except in the experiments to determine optimal internal temperature (Chapter 5) joints were cooked to an internal temperature of 75°C. Grant recorders were used to monitor both air temperature and internal temperature in all experiments. Air temperature was adjusted when required using readings from the sensor rather than relying on an oven thermostat. The sensor was always located in the centre of the oven. In many experiments, a forced air convector oven was used to ensure greater uniformity of temperature throughout the cooking process. All the joints studied in Chapter 5 and in the Grass/Rape trials were cooked in this way.

Thermistors for monitoring internal temperature were always inserted in the same position. Minor adjustments were made, if required, in response to changes induced by the cooking process. Once an internal temperature of 70°C was reached, careful attention was required in that internal temperature subsequently rose by 1°C/min.

6. Post-Cooking Procedures

The required internal temperature attained, joints were reweighed hot immediately. Samples of fat and lean were removed as indicated in section 3 of this chapter for pH determinations and assessment of cooked aroma. Separate equipment was used for each joint. None of the/

the external browned area was used. Total weight losses for each joint were calculated and expressed as a percentage of raw on the bone weights. Drip losses - as determined by increased in weight of the standard roasting tins - were also expressed as a percentage of raw on the bone weight of joints. Subtraction of drip losses from total weight losses enabled evaporative losses to be calculated. pH values of the cooked slurry were determined by the use of a calibrated pH meter. Should there appear to have been no change or a fall in pH value as a result of the cooking process, the value was always checked.

Except in experiments described in Chapter 5, joints were cooled and refrigerated ($5^{o}C$) prior to slicing. Joints were sliced with a Moulinex electric carver. Slices were converted to small, uniform $10mm^2$ portions for presentation to tasters. This involved removal of visible fat, the outer border of the meat and any areas of obvious gristle and connective tissue. This practice allowed samples to have, as far as possible, uniformity of texture and flavour. Whilst some would argue that removal of visible fat could affect flavour, this practice is now common and is indeed a dietary recommendation (Eating for Health 1978). Previous studies including those of Patterson (1975) and more recently Mottram and Edwards (1983) suggest that provided fatty tissue is not removed before cooking, its contribution to post-cooking flavour will not be affected.

In any case, for comparative assessment of flavour and aroma it is important that other product attributes are as nearly identical as possible. This minimises the distracting effect of 'clues' on subjects' performance.

7. Presentation of Samples to Tasters for Flavour Assessment

Samples were presented to tasters on white 175mm disposable plates./

plates. Samples were identified by the use of a random number three digit code. Examples are shown (Plates 5 and 6 of the Appendix). Reserve supplies of meats were always available to ensure that any one of the samples in the paired comparisons or triangle tests did not 'prompt' responses by differing in quantity! Although recording sheets used in trials were the same, as indicated in Chapter 3, actual presentations varied to minimise positional bias.

Participants were provided with water and cream crackers to act as palate cleaners at tasting sessions. Since a report on aftertaste was not required, it was emphasised that after flavour assessments, meats need not be swallowed. Paper towels allowed for them to be discarded. Problems of alteration in response as a result of post-ingestional satiety were thus avoided. A laissez faire approach was however used to encourage co-operation. Retasting was allowed.

Every effort was made to ensure that the atmosphere in the laboratory was calm, purposeful but relaxed. Subjects were encouraged to work at their chosen pace. They appeared to work conscientiously and independently and with willingness despite the test periods often being during a lunch break. (The latter is well recognised as being far from ideal for testing procedures of this type). In later trials, separate booths in a specialist sensory appraisal room were used. Orange lighting masked any slight difference between samples.

In such experiments, the most meticulous care and attention to detail are required of the experimenter if diferences in appearance and texture are to be minimised. Both can influence flavour perception. A subject's response to an unexpected, albeit tiny, portion of gristle could exert strong influence on preference. Most of/

of the preparations were therefore carried out by the experimenter with minimal technical assistance.

Subjects were asked which sample of a pair they preferred on the basis of flavour. Triangle tests were also carried out to identify the odd sample, again on the basis of flavour. The design of the experiments is indicated in Table 6.2 (Chapter 6).

8. Presentation of Samples for Assessment of Aroma

In these trials, uniformity of appearance and texture posed few problems. In Chapter 3, it has been noted that differences in aroma were expected between test and control samples and that randomised presentations AAB, ABA and BAA were used with B being the test sample. Raw samples were assessed in trials 2, 3 and 5 and cooked samples in trials 1, 4 and 6. It was emphasised that trials could be carried out in any order but a specially designated area of the laboratory should be used for each trial.

Separate tiles, forceps and scalpels were used in the preparation of the six A and three B samples for each trial. Samples were transferred to Quickfit 50.00ml boiling tubes fitted with ground glass stoppers. Each tube and stopper was marked a day in advance with the appropriate three digit random number code using a Pentel marker. The three stoppered tubes for each trial were clipped together with an elastic band and immersed in an aluminium beaker containing 100ml water. For 30 min. prior to testing, each of the six beakers was immersed in a thermostatically controlled water bath ($60^{o}C$) to promote release of 'volatiles'. Testers were requested to remove beakers for each trial from the waterbath and to make their judgement in the specially designated area of the laboratory. Trials could be carried out in any order and when convenient within a given tasting session.

After/

After removing the stoppers, contents of each of the three tubes were assessed for aroma. Recognition of the odd sample of the three was required. Subjects were encouraged to respond even if they considered that they were 'guessing'. Failure to respond was considered to be a failure to identify the odd sample of the trio. By randomising presentations it was hoped to minimise learning effects. After immersion in the waterbath at 60°C for 30 minutes there was no obvious difference between raw and cooked samples to guide testers in their decision. After assessment, beakers containing the tubes were returned to the waterbath for reheating. Additional samples were available on rare occasions when ground glass stoppers jammed or boiling tubes fractured. Subjects worked at their own pace. The importance of reaching independent conclusions was emphasised. Positioning of designated areas of the laboratory made observation of facial expressions of other participants difficult. Again, extreme care was required of the experimenter. Procedures outlined were inevitably messy and unobtrusive clearing of designated areas of the laboratory was essential to avoid distraction of participants.

When flavour (paired comparisons and triangle tests) and aroma (triangle tests) were completed, after checking, recording forms were submitted to QMC's Data Processing Centre for analysis with computer support. These procedures had been refined following experiments described in Chapter 5.

9. Selection of Tasters

It had always been the original intention to use only tasters selected on the basis of preliminary screening tests. Individuals differ in sensitivity, interest, motivation and ability to judge differences. It was recognised that results of trained or experienced panels/

panels may not resemble those of consumer studies and that the desirable pretesting of the food to be assessed was impracticable. Aptitude for flavour assessment varies between individuals, between products and at different times for the same individuals and products. Some sacrifice of precision was required to save time and expense. Various screening procedures are used but possibly the selection of tasters on the basis of performance as described in the Deoiled Herring Silage trials (Related Studies) is the most practicable in the College situation. However, many participants in subsequent trials took part in the July 1977 experiments with lamb chops and were thus aware of what was required of them.

Spencer (1972), himself trained at Arthur D Little Research Co., selected panel members on their ability to identify primary tastes in aqueous solutions, performance in ranking sucrose solutions of differing concentration and in odour recognition tests. The extent to which such selection procedures are reflected in superior performance in actual tests is difficult to assess and Kramer et al (1961) considered it unwise to reject subjects on the basis of a single screening test. Spencer tested 147 people. Only 73 were selected on their ability to identify primary tastes of whom 67 took part in ranking tests. A further five were rejected. Of the 57 participants in the odour recognition tests only 43 remained of whom only 30 were subsequently trained as panel members. This represents a success rate of only 20%. Twentyone of the subjects who took part in trials described in Chapters 5, 6 and 8 had carried out similar testing procedures. A specimen recording form is included in the Appendix (Table 4.3). Eleven identified the primary tastes correctly. Two transposed tap water and the bitter stimulus. Six transposed sour and bitter stimuli. These are both very common/

common errors. Participants would normally be retested. The two others tested did not respond to the salt stimulus.

Twelve subjects ranked the 8, 10 and 12% sucrose solutions correctly. This is a very demanding procedure. In odour recognition tests, the experimenter awards scores for each description. A score of five is awarded for a named chemical, four for a colloquial equivalent through to one for any response which could be considered valid by the subject. To be accepted for training a score of 60-70% is generally required. Scores were very high. Three of the 21 were laboratory technicians. Two were nurses. Effects on scores are indicated in Table 4.1.

Table 4.1 Odour Recognition Scores

		N	$\bar{X}$	SD
1.	Total - all subjects	21	70.4	10.6
2.	Total - omitting laboratory technicians	18	68.1	8.4
3.	As 2 - omitting nurses	16	70.3	5.78

One technician scored 93%. The two nurses scored 50% and 51%.

Results of these tests were not subsequently linked to performance in lamb trials. As is indicated in Chapter 5, participants frequently changed and the nurses mentioned above had left College before the main trialsstarted. There would have been too few subjects to achieve valid results. However, they are of interest and tend to confirm the observation in other chapters that QMC students and staff perform well in sensory testing procedures.

Most workers now agree that no consistent relation between taste acuity and judges reponses can be demonstrated. However, although it is only one factor in discriminatory ability, it is reasonable/

reasonable to expect tasters to identify the four primary tastes. Girardot (1952) indicated that the Quartermaster group rejected those who could not do so. Whilst such screening procedures avoids including subjects of poor ability it does not identify in advance those who lose interest as experimental trials continue. Mackey and Jones (1954) showed that high sensitivity did not correlate well with ability to arrange foods in order of stimulus concentration. Sequential performance in triangle tests is used by many experimenters (Winger and Pope 1981). In most situations, Amerine et al (1965) comment that elaborate pre-testing procedures are probably unnecessary.

A majority of panel members had either training or experience in the use of sensory appraisal techniques. In requesting volunteers for new stages of the trials, it was not considered desirable to reject those who were willing to participate. Those whose results showed little aptitude could be dropped from the trials without their knowledge. In retrospect, those selected as panel members might have considered themselves an elite group which could have increased motivation and have stimulated others to apply!

10. Factors Affecting Sensitivity of Panel Members

Good health is important. Pain, discomfort or tiredness interfere with judgements. Sensitivity declines after the age of 50 but experience and training may compensate. In women, the stage of the menstrual cycle may be of importance. It is claimed that odour sensitivity may be greater premenstrually but a survey carried out by the author suggests that a great deal of careful experimentation would be required to corroborate this view. Contrary to popular opinion, smokers and non-smokers perform equally well (Spencer 1972). Requiring tasters to abstain from smoking for 30 minutes before tasting sessions has more to do/

do with the distracting effect of tobacco on non-smokers than smokers' own performance. Cosmetics and highly perfumed soaps have similar distracting effects.

Interest, motivation, competition and knowledge of test results are well recognised psychological factors affecting sensitivity. It is often difficult to indicate individuals' results without affecting subsequent test procedures or discouraging the poor performer, but attempts were made to keep participants informed of experimental results during trials. Judges may become discouraged and lose interest when little difference is detected between samples. This could perhaps explain the falling off in attendance during the experiments described in Chapters 5 and 6 particularly. Subjects apparently like to find differences between samples. In using technicians and employees, interest cannot be assumed. None of these generally accepted beliefs seems to have been tested experimentally possibly for reasons of time and cost. Some examples of the experimenter's contact with tasters are included in the Appendix. (Tables 4.4 to 4.9).

11. Panel Size

In the guide prepared by the Committee on Sensory Evaluation of the Institute of Food Technologists (Anon. 1964), recommendations are for three to ten trained, eight to 25 semi-trained or over 80 untrained judges. BS 5929 also advises on panel sizes. Panels for preference testing which is so highly subjective should be larger than those for difference testing. Amerine et al (1965) recommend 10-20 judges with three or four replications per judge. This recommendation was followed wherever possible in the present studies but, in the event, panel size was ultimately determined by availability and willingness to attend tasting sessions.

Discussion/

Discussion with students who had taken part in previous trials suggested that whilst out-of-pocket expenses should be reimbursed, payment could induce volunteering for the 'wrong' reasons and would be extremely expensive.

Volunteers were thus requested from QMC staff and students. However, few people seemed to read notices until trials were completed and individual notices fared little better. In a Scottish Central Institution where teaching is regarded as the main commitment, staff are not always willing to release individuals or groups for tasting sessions which are not seen to be directly related to the learning process. Others feel that students should not be 'used' in this way. It was not found that there was a large pool of students to act as tasters, nor that because large numbers of them and those who teach them are taking part in food orientated courses commitment to taste panels would be a priority.

The problem is not unique to QMC. Of 29 ESCA volunteers, only six remained latterly. In research institutes or in the food industry, it is recognised that taste panel work is a responsibility of one's post. In this situation, availability is the main criterion for membership but even then, day release classes, attendance at meetings and absences on business still pose problems of providing orthogonal groups.

It must thus be accepted that in sensory appraisal tests, panel sizes are often too small to satisfy the statistician and that this is a problem almost inherent to the technique. The experimenter cannot coerce, since co-operation and motivation of tasters are essential. It is difficult to explain that attendance is crucial to good experimental design, particularly as reasons, apart from forgetting, are usually/

usually valid. In any case, with commodities such as meats, limited yield and cost are additional constraints.

However, set in context, when tasting sessions form part of a teaching programme at either QMC or ESCA panel sizes are adequate and members have been shown to perform well. Much of the present work was carried out in this way.

12. Testing Environment and Presentation of Samples

Regular sessions, quietness, comfort, orderliness and attractive presentation of samples are essential prerequisites in sensory appraisal testing. Uniformity of presentation to achieve comparability between tasters and from panel to panel is necessary. Temperature, humidity, rate of air flow and air purity are important and are more readily controlled in specialist accommodation such as that available at QMC. Independent judgements are essential so that separate booths are of great assistance as are neutral background of ceiling, walls and floor. It is desirable that coloured lighting can be used to disguise slight differences in appearance of foods which could create flavour anticipation.

Optimal timing is controversial and may be influenced by the product being tested. Mid and late morning and afternoon are generally most preferred with panels immediately after meals least preferred. Christie (1962) noted that meat tasters performed slightly better when hungry. For reasons indicated elsewhere in this study, there was no choice of time. Sessions took place whenever participants were available.

For reasons of economy of panel time, the maximum number of samples compatible with avoiding sensory fatigue should be assessed at a session. If samples are tasted and discarded and palate cleaners are used, the onset/

onset of sensory fatigue can be delayed. Thus it proved possible to present four replicates instead of two in triangle tests. When surveyed, the majority of participants judged the number of samples to be 'about right'.

Previous experimentation and results described in Chapter 5 indicate that the procedures devised for the present trials were satisfactory. In particular, tasters were assessing lamb samples as they are traditionally prepared in the United Kingdom.

CHAPTER 5

Summary - Determining the Optimal Internal Temperature for Lamb Joints.

Gigots and loins of standard anatomical location were cooked under controlled conditions to four different internal temperatures. Hedonic rating scales were used to assess optimal internal temperatures, if left and right joints from the same carcass differed and to assess tasters' consistency in duplicated tests and to compare their acuity when samples were tasted either hot or cold. It was possible to achieve sufficiently meaningful results to make recommendations for future experimental investigations.

Determining Optimal Internal Temperatures for Lamb Joints

1. Since the AAHE recommendations of 1971, it has been considered that internal temperatures of 60°C, 71°C and 77°C will normally produce beef roasts defined as rare, medium rare and well done respectively. It is recognised that sex, age of animal at slaughter, extent of carcass ripening and cooking temperature may alter the temperature of colour changes and hence assessment of doneness. In the present experiments these factors were controlled. Although there have been many studies of beef, fewer appear to have been carried out with lamb in the United Kingdom. Rhodes (1976) in reporting on lamb joints from sheep fed lucerne treated with formaldehyde or glutaraldehyde, indicated that an internal temperature of 80° was selected at MRI. An oven temperature of 150°C (300°F approximately) was used. The AAHE recommendations for lamb are that an internal temperature of 65°C should be attained. Normally the U.S.Department of Agriculture recommends a cooking temperature of 325°F (163°C) to combine maximum yield, juiciness and tenderness in roasted meats.

The/

The time/temperature relationship makes careful control of cooking temperature essential (Chapter 4). However an oven temperature of 177°C (350°F approximately) used in the present experiments is considered to correspond more closely to culinary practice in the United Kingdom. It is accepted that there is variation in roasting techniques. The experimenter considered that there were indications that individuals' preference for 'doneness' even in lamb show variation, although possibly to a lesser extent than for beef. It was thus decided to cook joints under standard conditions to four different internal temperatures - 70°C, 75°C, 80°C and 85°C. Left hand side (LHS) and right hand side (RHS) gigot (G) and loin (L) roasts were cooked. Experiments were carried out as follows: (Table 5.1)

Table 5.1 Allocation of Joints to Internal Temperatures

Joints*	Internal temperature attained °C
A, E	70
B, F	75
C, G	80
D, H	85

*Obtained from the carcasses of eight Scottish Blackface wethers of comparable weight and feeding regime from the ESCA farms.

It can thus be seen that the four joints obtained from carcasses A and E were cooked to an internal temperature of 70°C. A similar practice was followed for the four joints from the remaining three pairs of carcasses cooked to internal temperatures of 75°C, 80°C and 85°C respectively. Apart from loin roasts A to D, LHS and RHS, all trials, unknown to tasters, were duplicated. Presentation of samples to them was as indicated in Appendix Table 5.1. A seven point hedonic rating scale was used (Appendix Table 5.2). Despite its apparent simplicity the limitations of assessing meats by hedonic scales is recognised particularly/

particularly when the number of participants was not always greater than 20. This qualification is reinforced by the analysis of tasters' scores in duplicated tests. Consistency was on the whole poor. Lack of consistency inevitably restricts the conclusions to be drawn from such experiments.

Results were analysed using Friedman's Two-Way Analysis of Variance. Reasons for selecting this non-parametric statistical technique are indicated in the section on Statistical Procedures in Chapter 3. Since the value of N always exceeded 9 and K was 4, standard χ^2 tables were used to determine significant levels of the Chi R square values. Column totals i.e. the ranked scores for lamb of a particular internal temperature provide some indication of the differences in attitude towards each sample. Sign tests were used to provide more precise information. Tables of the most frequently ranked scores, median values and 'like' responses (i.e. scores exceeding 4) were constructed since the only result of statistical significance in the whole series of trials ($p<0.05$) was for a loin roast cooked to an internal temperature of $75^{o}C$. Tasters' results in duplicated tests were correlated and Spearman Rho values calculated. In analysing the results, interest was focussed on the following:

(i) Is the optimal internal temperature for a lean joint (gigot) the same as for a less lean joint (loin)?

(ii) Do LHS and RHS gigots and loins from the same carcass differ?

(iii) How consistent are tasters when duplicated samples are presented? (A variation of plus or minus one is expected when hedonic rating scales are used)

(iv)/

(iv) Is tasters' acuity greater when meat samples are tasted hot rather than cold?

Table 5.2 The Experimental Design - January 1978

File LA*	Subjects (N)	Joints	Cut	Side	Temperature
B)	27	A-D	G	R	C
C)	28	"	"	R	C
E)	11	"	"	L	H
G)	11	"	"	L	H
I)	12	E-H	"	L	C
J)	12	"	"	L	C
F)	12	"	"	R	H
H)	12	"	"	R	H
A	29	A-D	L	Le	C
D	28	"	"	R	H
M)	15	E-H	"	R	C
N)	16	"	"	R	C
K)	16	"	"	Le	H
L)	16	"	"	Le	H

where G = gigot L = loin C = cold H = hot
Le = left R = right

Brackets indicate duplicate samples presented to tasters on the same occasion.

In addition to monitoring tasters' performance it was hoped to determine if acuity varied according to whether samples were tasted hot or cold. Thus the following comparisons were made using computer files LA*. (Tables 5.3 and 5.4)

Table 5.3 To Determine Tasters' Consistency in Duplicate Tests.

Gigots	N	Loins	N
B v C	27,28	M v N	15,16
E v G	11,11	K v L	16,16
I v J	12,12		
F v H	12,12		

Table 5.4/

Table 5.4 To Compare Tasters Consistency/Acuity when Samples are Tasted Cold or Hot

Gigots	N	Loins	N
B v E	27, 11	A v D	29, 28
B v G	27, 11	M v K	15, 16
C v E	28, 11	M v L	15, 16
C v G	28, 11	N v K	16, 16
I v F	12, 12	N v L	16, 16
J v F	12, 12		
I v H	12, 12		
J v H	12, 12		

In making this second group of comparisons, B (the sample tasted cold) is compared with both E and G the comparable samples tasted hot. A similar comparison is made with the duplicate of B, C which is also compared with E and G. It will be noted that in comparing cold and hot samples that they were from left and right sides of the same carcass of standard anatomical location. It was thus assumed that left and right sides would respond comparably to the cooking process. Since, despite differences in initial pH values of left and right joints of standard anatomical location, the change in pH induced during the cooking process shows no difference of statistical significance, this assumption was subsequently shown to be justified. The crossover between left and right sides between carcasses A to D and E to H should also be noted.

An example of a computer printout for Friedman's Two Way Analysis of Variance and of the comparisons made is included in the Appendix (Table 5.4). Sign Tests were not - as they are at present - carried out with computer support. The printout also indicates means and standard deviations. At the time, they were included for interest only but subsequently the medians were calculated and used in the analysis of the data. This practice/

practice is of greater validity (See discussion in Chapters 3 and 12).

2. Results of the Trials

These are presented in Table 5.5 and are in the same sequence as in Table 5.6 of this chapter. Values of the Median and the Rank Column Totals, together with values of Chi R square are recorded. Results are considered to be of statistical significance if $p<0.05$ but for interest values of $p<0.10$ are presented since the practice is followed by some research workers. A more recent printout is also included (Appendix Table 5.5).

The table gives little indication of obvious preference for a particular internal temperature. There is only one statistical significance preference ($p<0.05$) for a loin roast cooked to an internal temperature of 75°C. The lowest rank column total is underscored for each trial. The most frequently occurring lowest value is to be found for roasts cooked to an internal temperature of 75°C.

The Sign test was used to compare each pair of values. Results are tabulated in Table 5.6. The similarity of the preferences for all samples is again emphasised. Since the direction of difference was not specified in advance, the test is two-tailed. Hence probability values in the table must be doubled. (Values were obtained from Table M Haber & Runyon 1977). In trial B, an internal temperature of 70°C was rated unfavourably in comparison with 75°C and 85°C, whilst in trial C, 80°C was unfavourably rated as compared to 85°C. In trial E, there was a statistically significant preference ($p<0.04$) for 75°C over 70°C. These were gigot roasts. For loins, there was a statistically significant preference ($p<0.01$) for 75°C over 70°C. On balance, examining the Rank Column Totals and by the use of the Sign Test/

Table 5.5 Medians and Rank Column Totals for Joints Cooked to Different Internal Temperatures

File LA	Cut	N	A (70°C)		B (75°C)		C (80°C)		D (85°C)		Chi R Square	Probability *
			Median	CT	Median	CT	Median	CT	Median	CT		
B	G	27	3	83.5	3	62.5	3	64	3	60	7.77	<0.10
C	"	28	3	72	2	66.5	2	76	3	66.5	1.55	NS
E	"	11	5	34.5	3	20.5	4	28	4	27	5.37	NS
G	"	11	5	32	2	20.5	4	29	3	28.5	3.95	NS
I	"	12	4	27.5	4.5	28.5	4	29	4.5	35	1.73	NS
J	"	12	3	29.5	4	32	3.5	26	4	32.5	1.33	NS
F	"	12	4	32.5	3	25	3.5	33.5	3	29	2.23	NS
H	"	12	3.5	24.5	3	28	4.5	37.5	3	30	4.53	NS
A	L	29	4	87.5	3	61	3	67	4	74.5	8.10	<0.05
D	"	28	4	72	3	61	3	69.5	3	77.5	3.03	NS
M	"	15	4	44	4	34	4	38.5	3	33.5	2.86	NS
N	"	16	3.5	39	3.5	39.5	4	42	4	39.5	0.21	NS
K	"	16	3	34.5	3	40	4	43	3.5	42.5	1.71	NS
L	"	16	2	32.5	2	41.5	3	46	2.5	40	3.54	NS

* df = 3

CT = Ranked Column Total

Table 5.6 Sign Tests - Probabilities. (Allowance has been made for a two-tailed test)

File LA*	Cut	N	AB	AC	AD	BC	BD	CD	CTA	CTB	CTC	CTD
B	G	27	032	054	004	500	500	500	83.5	62.5	64	60
C	"	28	271	212	202	425	584	039	72	66.5	76	66.5
E	"	11	020	254	145	363	090	637	34.5	20.5	28	27
G	"	11	062	062	109	500	254	500	32	20.5	29	28.5
I	"	12	500	623	145	500	227	254	27.5	28.5	29	35
J	"	12	500	500	500	377	637	227	29.5	32	26	26.5
F	"	12	377	500	377	145	145	145	32.5	25	33.5	29
H	"	12	254	033	500	090	363	254	24.5	28	37.5	30
A	L	29	005	032	143	252	212	332	87.5	61	67	74.5
D	"	28	416	416	584	252	054	164	72	61	69.5	77.5
M	"	15	059	500	113	291	500	387	44	34	38.5	33.5
N	"	16	613	291	500	291	377	402	39	39.5	42	39.5
K	"	16	291	212	194	291	605	500	34.5	40	43	42.5
L	"	16	212	059	377	387	500	254	32.5	41.5	46	40

Decimal point for .XXX omitted. Results of statistical significance underscored.

Abbreviation CT as in Table 5.5.

Test there are indications that an internal temperature of 75°C is preferred. Since the decision was critical for future experiments, the median values and the "like" responses were studied in more detail.

The response patterns of the tasters for gigots and loins is indicated in Table 5.5. The bracketed trials are the duplicate samples tasted on the same day. It can be seen that the frequency distribution of responses in duplicate trials is not in close agreement. To overcome this difficulty, median values were studied as were total 'like' responses i.e. those responses exceeding a value of 4.

Table 5.7 Median Values of Gigots

File	T°	A (70°C)	B(75°C)	C(80°C)	D(85°C)
B)	C	3	3	3	3
C)	"	3	2	2	3
E)	H	5	3	4	4
G)	"	5	2	4	3
F)	H	4	3	3.5	3
H)	"	3.5	3	4.5	3
I)	C	4	4.5	4	4.5
J)	"	3	4	3.5	4
Sum		30.5	24.5	28.5	27.5
Ranked Column Total		22.5	16	20	20.5

The median values for the roasts at each internal temperature were totalled to give an indication of preference. There is little difference between samples with B being most and A least preferred. If the medians for the 4 joints in each trial are ranked, the ranked column totals confirm the observation that B is most preferred and A least.

The relative position is slightly altered if total responses exceeding/

exceeding 4 i.e. the "like" responses are studied (Table 5.8). The order of preference B, D, A, C when both total "like" responses and rank column totals are studied. It will be noted that low values of the median indicate a higher preference for samples whereas the converse applies to the "like" responses.

Table 5.8 Gigots - Numbers of "Like" Responses i.e. Scores Exceeding Four

File	T°	N	A	B	C	D
B)	C	28	16	18	15	18
C)	"	"	21	21	21	18
E)	H	11	3	8	5	3
G)	"	"	4	6	5	6
F)	H	12	5	7	6	7
H)	"	"	6	8	4	9
I)	C	12	6	5	5	5
J)	"	"	8	5	6	5
Sum			69	78	67	71
Ranked Column Totals (Order of Preference)			21.5	15.5	23	19.5

It is however clear that there is very little difference in response to samples - possibly because as one taster remarked "the test is very difficult because they all taste so good". In all tests B is the preferred sample.

Loins were studied in the same way. Results are presented in Table 5.9 and 5.10. They differ considerably from those for gigots

Table 5.9 Median Values of Loins

File	T°	A(70°C)	B(75°C)	C(80°C)	D(85°C)
A	C	4	3	3	4
D	C	4	3	3	3
K)	H	3	3	4	3.5
L)	H	2	2	3	2.5
M)	C	4	4	4	3
N)	C	3.5	3.5	4	4
Sum		20.5	18.5	21	20
Ranked Column Total		15	11	18	16

Table 5.10 Loins - Numbers of Like Responses i.e. Scores Exceeding Four

File	T°	N	A	B	C	D
A	C	29	8	17	15	14
D	C	27	10	18	17	18
K)	H	16	12	9	7	8
L)	H	"	12	13	10	10
M)	C	15	7	6	7	8
N)	C	16	9	8	6	7
Sum			58	71	62	65
Ranked Column Totals (Order of Preference)			14.5	10	20.5	15

The order of preference for loins as judged by median values are B, D, A, C when totalled and B, A, D, C when ranked column totals are studied. When scores for values exceeding 4 are totalled, the order of preferences is B, D, C, A. Corresponding preferences from the ranked column totals are B, A, D, C. Thus for loins, B is the preferred sample. It will be noted that results for the median tests are very close for all samples but that there are rather larger differences between total "like" scores.

The median scores for hot and cold gigots and loins were considered separately. They were classed in order of preference. Table 5.11 summarises these results.

Table 5.11 Preferences for Samples A, B, C and D Based on Ranked Median Values and Like Scores

Cut	T°	A	B	C	D	
G	C	2	3	1	4	
"	H	4	1	3	2	Medians
L	C	1.5	1.5	3.5	3.5	
"	H	2.5	1	4	2.5	
G	C	1	2	3	4	
"	H	1.5	4	3	1.5	"Like"
L	C	4	1	3	2	Responses
"	H	3.5	1	3.5	2	

For loin samples, B is preferred both cold and hot. Results are more uniform. This is perhaps to be expected since loin roasts have been shown to be more homogeneous than gigots elsewhere is this study. For cold gigots A is the preferred sample with B, C and D resembling each other but B is preferred when gigots are tasted hot. When the medians and "like" values are compared there is no very great difference between them but, again, on balance 75°C (B) seems to be more generally preferred.

For interest a summary of the response patterns is shown in Table 5.12. More detailed information is provided in the Appendix (Table 5.1).

Referring again to results relating to the values of Chi R square, the rank column totals for each of the internal temperatures (70°C, 75°C, 80°C and 85°C) were themselves ranked. Ties are indicated. These rank orders are indicated in Table 5.13. When they are studied, it suggests that the most preferred internal temperature for both gigots and loins is 75°C.

Table 5.12 A Summary of the Response Patterns

File	Cut	T°	N	Median Values			
				A(70°C)	B(75°C)	C(80°C)	D(85°C)
B)	G	C	28	3	3	3	3
C)	"	C	"	3	2	2	3
E)	"	H	11	5	3	4	4
G)	"	H	"	5	2	4	3
F)	"	H	12	4	3	3.5	3
H)	"	H	"	3.5	3	4.5	3
I)	"	C	12	4	4.5	4	4.5
J)	"	C	"	3	4	3.5	4
A	L	C	29	4	3	3	4
D	"	C	27	4	3	3	3
K)	"	H	16	3	3	4	3.5
L)	"	H	"	2	2	3	2.5
M)	"	C	15	4	4	4	3
N)	"	C	16	3.5	3.5	4	4

Table 5.13 Most Frequently Occurring Preferences as Judged by Rank Column Totals

File LA*	Cut	70°C	75°C	80°C	85°C
B	G	4	2	3	1
C	"	3	1.5	4	1.5
E	"	4	1	3	2
G	"	4	1	3	2
F	"	4	1	3	2
H	"	1	2	4	3
I	"	1	2	3	4
J	"	2	3	1	4
Sum		23	13.5	24	19.5
A	L	4	1	2	3
D	"	3	1	2	4
K	"	1	2	4	3
L	"	1	3	4	2
M	"	4	2	3	1
N	"	1	2.5	4	2.5
Sum		14	11.5	19	15.5

Although preference for 75°C is evident gigots and loins cooked to 85°C were also well liked. However, as internal temperature rises, weight losses increase and juiciness decreases. Although in Chapter 10 it is demonstrated that weight losses are variable, in this series of trials they followed the theoretical pattern as Table 5.14 indicates.

Table 5.14 Total Weight Losses with Increasing Internal Temperature

			% Losses				Rank Losses			
Joints	Cut	Side	70°C	75°C	80°C	85°C	70°C	75°C	80°C	85°C
A-D	G	Le	25.8	28.6	28.9	42.3	1	2	3	4
"		R	24.8	32.6	38.4	41.2	1	2	3	4
E-H		Le	20.6	31.4	34.5	42.5	1	2	3	4
		R	19.8	21.2	33.7	36.0	1	2	3	4
A-D	L	Le	12.3	18.5	8.2	14.3	1	4	2	3
	"	R	19.5	16.8	21.8	16.7	3	2	4	1
E-H	"	Le	17.0	24.5	30.5	29.9	1	2	4	3

Thus/

Thus, preference apart, there would be more meat available for tasting purposes if the internal temperature of 75°C were to be selected.

Spearman Rho Values, that is rank order correlation coefficients were calculated for each participant's responses to the four samples both for duplicated trials and for trials intended to compare acuity when samples were tasted either hot or cold. Tables are available which enable the testing of the null hypothesis that the ranks are independent for relatively small values of N. Because there are only four ranks in each test in these experiments, the calculated value of r must be 1.000 to reach a level of significance of $p<0.004$. By the nature of the Hedonic Rating Scale, subjects, as indicated, may score a particular sample plus or minus one of the allocated score. Although this appears to be imprecise, as indicated by previous discussions in this chapter, meaningful tendencies become evident provided there are sufficient participants in a given trial. However when examining the Spearman Rho values, a very wide scatter between values 1.000 and -1.000 is apparent.

Considerable study has been made of this data. Some examples of the problems encountered are provided. If values are dichotomised so that values of Spearman Rho from 1.000 to zero are grouped and compared with minus values, the following results are achieved. The reader is referred again to Table 5.3.

Table 5.15 Duplicated Trials to Assess Tasters' Consistency

Cut	T°	Files LA*	1.000-0	Minus Values	N
L	C	M v N	13	2	15
"	H	K v L	8	8	16
G	C	B v C	1	10	27
"	"	I v J	6	6	12
"	H	E v G	10	1	11
"	"	F v H	8	4	12

For the loin samples the calculated value of χ^2 is 5.32 and for the gigots, cold and hot 0.50 and 1.06 respectively. Thus for loins p is <0.05. There would be limited value in pooling results for gigots, but the calculated value of χ^2 is 3.48 corresponding to $p<0.10 >0.05$.

Comparisons are made more difficult in that subjects did not attend sufficient of the sessions to allow meaningful comparisons to be made. This problem is highlighted in Table 5.16 where orthogonal samples are few.

Table 5.16 Duplicated Samples - Spearman Rho Values

Subject	Cut	Cold (MN)	Hot (KL)
7	L	0.6	0.9
14	"	0.3	0.8
19	"	0.6	0.2
24	"	0.8	0.3
25	"	1.0	0.7
		Cold (BC)	Hot (GH)
12	G	0.6	1.0
17	"	0.2	0.3
24	"	0.9	0
26	"	0.8	0
32	"	0.1	0.7
34	"	0.1	0

The same problems of few comparisons and great variability in rank correlation coefficients when assessments of joints cold and hot are compared. Tables 5.2 and 5.3 of the Appendix indicate why it was necessary to reach decisions regarding testing procedures in the absence of correlation coefficients. The distribution of the correlation coefficients is indicated in the Appendix (Table 5.6).

The difficulty in the use of descriptive scoring techniques - in this case a more structured hedonic rating scale - is further confirmed when results of tasting sessions carried out by ESCA BSc students of Agriculture are studied./

studied. Lamb joints were cooked to internal temperatures of 70°C/ 75°C (L samples) and 80°C (H samples). Tenderness, juiciness and flavour were allocated scores of 1, 3, 5, 7 and 9 according to descriptions of these characteristics which were considered most appropriate. Although on this nine point scale intermediate values could have been used, they seldom were. Thus, for the majority of tasters, the scale was effectively a five-point scale. The instructions for the test are set out in Table 5.7 of the Appendix.

The results of the tests are presented in Table 5.17. Results for 1977 and 1979 are considered separately from the others. For these two years, results were not presented as arithmetic means, standard deviations with quoted values of χ^2. Instead, scores were ranked, rank column totals and the Chi R square statistic calculated. At this stage, it is not possible to recalculate the experimental figures. Recording forms were returned to the students following the testing procedures and the collating forms are no longer held in the Data Processing Centre files.

It will be noted from Table 5.7 of the Appendix that Tenderness, Juiciness and Flavour were assessed for each of the L and H samples. Although the present study is concentrated on lamb flavour, tenderness and juiciness may exert influence on flavour perception.

From Table 5.17, it will be noted that, despite differences in internal temperature of 5°C or 10°C, there was little difference in the arithmetic means or rank column totals in relation both to tenderness and flavour. No statistically significant difference as indicated by χ^2 and Chi R square values was thus established between L and H samples. When juiciness is considered, in 1977 and 1979, statistically significant differences ($p<0.05$ and $p<0.001$) were demonstrated with the L samples - as/

Table 5.17 Descriptive Scoring of Tenderness, Juiciness and Flavour of Lamb Samples Cooked to Different Internal Temperatures

Year	Tenderness				Juiciness				Flavour			
	70°/75°C		80°C		70°/75°C		80°C		70°/75°C		80°C	
	AM	SD	AM	SD	AM	SD	AM	SD	AM	SD	AM	SD
1976	4.16	1.52	4.16	1.44	4.39	1.67	4.84	1.64	4.11	1.84	4.11	1.68
1980	3.62	1.64	3.58	1.70	4.58	1.67	4.88	1.86	4.03	1.53	4.18	1.76
1981	4.10	1.53	4.62	1.82	4.26	1.66	4.06	1.75	4.72	1.44	4.50	1.79
1983	4.55	1.69	4.52	1.65	5.42	2.00	4.74	1.77	4.68	1.64	4.81	1.96

	CTL	CTH	$ChiR^2$	CTL	CTH	$ChiR^2$	CTL	CTH	$ChiR^2$
1977	77.5	75.5	0.08	65.5	87.5	9.4*	73.5	79.5	0.07
1979	86	82	0.29	62.5	105.5	33.02***	82	86	0.29

where AM = arithmetic mean, SD = Standard Deviation, CTL and CTH = ranked column totals L and H respectively, * = $p<0.05$ and *** = $p<0.001$

as expected - judged the more juicy. In 1976 and 1980 the arithmetic means indicated a similar tendency. The findings in 1981 and 1983 are unexpected in that the H samples were considered to be juicier. Values of the standard deviation were however high. It should also be remembered that this test formed part of an educational exercise in each of the years. It was used to demonstrate that hedonic rating scales are of different format. Had this test been considered the most suitable, it would have been used by the experimenter in the trials described earlier in this chapter.

Since good performance by these students has so often been demonstrated in sensory appraisal tests during the seven years of experimental programmes, it can reasonably be concluded that differences in internal temperature of this order do not produce flavour differences sufficient to influence preference. In most experiments, variation in internal temperature of plus or minus 10°C are not readily detected except for alteration in juiciness.

3. Discussion and Conclusions

Although the use of hedonic rating scales to assess preference seemed initially to be more attractive than an extensive series of paired comparison (preference) tests, it is evident that, because of participants liked most of the samples, no clear cut preferences emerged from the trials. Study of the data, particularly in view of tasters' inconsistency, indicates the following:

1. There is no evidence to suggest different optimal internal temperatures for gigots and loins. The optimum for both appears to be 75°C.
2. There is no evidence that right and left joints from the same carcass differ in flavour.

3./

3. There is no evidence that acuity is greater when meats are tasted hot rather than cold.

It was thus decided that all gigot and loin roasts should reach an internal cooking temperature of 75°C. To ensure uniformity of presentation throughout tasting sessions and for practical reasons, meat samples should be presented cold. Although it could be argued that roasted meats are more usually eaten hot and thus motivation of tasters would be increased, there are logistical problems of maintaining them at the required temperatures throughout the tasting session. Although highly desirable, a single presentation to assembled tasters is not feasible in the College situation. These constraints emphasise the practicality of cold tasting. This was agreed by J.M.Harries (personal communication) and in the paper by Harries et al (1956).

In conclusion, since future studies were intended to demonstrate differences between test and control samples, the use of hedonic rating scales was inappropriate. Different sensory appraisal techniques and experimental design were required.

CHAPTER 6

Summary - Grass versus Rape: First Series of Trials

The rationale of the experimental design is explained. Triangle tests were used to detect possible differences in the flavour and aroma of roasted gigots and loins from grass and rape fed lambs. Paired comparison (preference) tests were also used. Analysis of the data by statistical procedures described in Chapter 3 indicated that there was no flavour difference between samples in this series of trials. It was thus decided both to repeat the trials and to use flavour profile analysis to confirm the present results. Tasters' acuity would be assessed by procedures described in Chapter 7.

1. Grass versus Rape. First Series of Trials

From Chapters 3 and 4 it is evident that there is wide variation in cooking procedures and sensory evaluation techniques between laboratories. Hence comparison of findings of one laboratory with another is difficult. A further complication arises in that statistical procedures are not standardised.

Much of the previous work reported relies heavily on the use of hedonic rating scales such as those described by Ford and Park (1980). Mean panel scores - scores being derived from a nominal scale used by trained panels - are used to indicate directional differences between samples. These scores are subsequently used to determine if differences between samples are of statistical significance.

In the present study, Chapter 5 indicates that the performance of untrained tasters - irrespective of experience - using hedonic rating scales is erratic rather than consistent. Findings obtained from untrained/

untrained tasters using hedonic rating scales should perhaps therefore be viewed with greater caution than has sometimes been the case.

It should be noted that the primary purpose of a hedonic rating scale is to indicate the attitude of tasters to a particular sample. In the untrained taster this attitude can be highly subjective as is indicated by subjects' bimodal response to the aroma of compounds such as phenylacetic acid (Land and Piggott 1981). In the analysis of data, nominal scales are subsequently converted to scores. At best, such scores are ordinal rather than interval in nature (Amerine et al 1965). Although the primary purpose of hedonic rating scales is to assess attitude to samples, if there are large and obvious differences between them, it becomes clear that particular samples are more or less acceptable to tasters and that differences thus exist between them. This practice has frequently been followed in previous studies designed to demonstrate if differences exist between samples. In addition, it seems to be assumed that either the midpoint of the scale or one of its extremes constitutes the desirable attribute - the former perhaps in relation to the aroma and flavour and the latter in relation to the tenderness of meats.

Because of the previously reported erratic performance of subjects in tasting tests (Chapter 5), it would require very large differences to exist between control and test samples for them to be highlighted by the use of this technique. It was thus decided to use specified difference tests to determine if differences did or did not exist between control and test samples. Further tests would subsequently be required to determine their nature and magnitude. Preference tests were also used. The ARC Statistician, Mr M M Franklin, confirmed that the/

the design of the proposed experiments was satisfactory. This design is indicated in Table 6.2 and a specimen recording form is included as Table 6.3 of the Appendix.

Because of the relatively small number of tasters at QMC, it was agreed that volunteers from ESCA should be recruited. To secure interest and motivation, joints from samples likely to differ in flavour would be compared. Thus the joints from the grass and rape fed animals were selected. The basis for selecting those feeding regimes for comparison was, as indicated, a series of experiments recorded by Park et al (1972). Tasting was divided between QMC and ESCA groups.

The animals used in the trials were Scottish Blackface wethers born April/May 1977. They remained with their mothers on the hill until August. After weaning, they were grazed on grass fields at lower altitudes until the start of the forage crops trial at the beginning of October when they were divided into three treatment groups receiving grass (A samples), rape (B samples) and stubble turnips (used for experiments described in Chapter 9). A small supplement of whole barley (approx. 135g) was introduced on 24.11.77 until slaughter on 6.12.77. Carcass weights of each group are indicated below.

Grass (A samples)		Rape (B samples)	
Animal No.	Carcass (kg)	Animal No.	Carcass (kg)
388	18.5	902	16.5
1240	17.5	1124	18.5
893	16.5	1464	15.5
914	16.0	194	18.5

Tasters always received a personal invitation to each session in addition to large notices reminding them posted prominently. As far as was consistent with avoiding influencing judgements, they were kept informed about the progress and results of tests as they proceeded (Appendix Tables 4.4 - 4.9).

In/

In the investigations described in this and subsequent chapters, the aim was to determine if differences in flavour did or did not exist in roasted lamb joints obtained from lambs on different feeding regimes. In view of previous reports of the effect of feeding regime on lamb flavour, it seemed probable that differences were likely to be established. Hence it was also considered important to establish if flavour preferences for lamb from different feeding regimes existed. The number of samples was small and the time scale short. Hence the recommendation of BS5929 (1980) that it is unwise to embark on an extensive series of preference studies prior to establishing statistically significant differences between samples was not followed. Such a recommendation, whilst theoretically desirable, may not be practicable in assessing meats and other commodities which show such inherent variation and limited storage life.

There are five recognised sensory evaluation procedures which can be used to determine if differences do or do not exist between samples. These are:

(a) Flavour Profile Analysis

(b) Paired Comparison (Difference) Tests

(c) Triangle Tests

(d) Duo - Trio

(e) Two in Five.

Each of these procedures is discussed in Chapter 3. Whilst having the undoubted advantage of immediate assessment of the magnitude and nature of differences between samples, flavour profile analysis is a specialised and time consuming procedure best reserved for the assessment of a small number of samples. The extensive training of the large number of potential panel members to participate in/

in this type of study was not deemed practicable.

The paired comparison (difference) test is one-tailed. It requires the experimenter to identify a particular attribute of the product being tested. For meats, it is suitable for assessing textural properties such as tenderness or juiciness, but for flavour a particular attribute is difficult to specify. 'Typical' lamb flavour is unlikely to have the same implications for every member of a taste panel. The choice of word implying taint or even strong flavour could create bias in attitude of panel members who may detect 'differences' between samples when none exist. For such reasons, particularly as the possible nature of flavour differences could not be specified, this test was considered unsuitable.

At the time the experimental programme was planned, the triangle test was considered to be the most appropriate for demonstrating if differences do or do not exist between samples. Muller (1977) believed that the position in the United Kingdom was likely to be similar to the United States where Brandt and Arnold (1977) had reported that it was the test most frequently used by food companies. Its main advantage is that the odd sample is identified correctly by chance alone on only one of three occasions. In the Duo-trio this possibility is one in two.

At the present time, comparative studies of the efficiency of triangle and duo-trio tests are in progress. An advantage of the latter is that more samples can be tasted at a given session thus making more economical use of tasters. However, it was not so highly regarded as the triangle test in 1977.

The two in five test, never as popular as the other two, was considered to be too complicated for use with untrained tasters who could/

could be working unsupervised. Thus a series of replicated paired comparison (preference) and triangle tests was planned. These would demonstrate if flavour preferences for a particular sample or flavour differences between samples existed.

During experiments with pork detectable difference in the aroma of subcutaneous fat from carcasses from test and control animals existed when the test animals had received 'deoiled herring silage' or 'spent brewers wash' in their diets. These differences were not present in cooked samples. In both experiments, these differences were not associated with flavour preferences. Substances responsible for these differences are likely to have volatilised or decomposed during the cooking process.

Thus in the trials to determine the effect of feeding regime on lamb flavour, in addition to carrying out tasting sessions using paired comparison (preference) and triangle tests on cooked lamb samples, triangle tests on the aroma of raw and cooked fat plus lean were carried out. Because it was assumed that differences would exist between samples, two control and one test sample were used in each triangle test. Each of the three possible presentations was used for both raw and cooked samples. It was considered that the resulting 18 assessments was the maximum which could be required of participants. The experimental design of the six presentations is indicated in Table 6.1.

Table 6.1 Design of Triangle Tests : Aroma of Raw and Cooked Fats

Batch	1	2	3	4	5	6
Presentation	BAA	AAB	ABA	AAB	BAA	ABA
R/C	C	R	R	C	R	C

where A = control (grass fed) sample
B = test (rape fed) sample
R = raw
C = cooked

The test procedures indicated in Chapter 4 were followed.

Since aroma is a major component of flavour, it was considered that these tests could be of value in identifying differences between samples.

In each of these experiments, performance of the subjects in the triangle tests on the raw and cooked samples was compared. Wilcoxon's Matched-Pairs Signed-Rank Test was used to test the hypothesis that subjects would identify differences between raw test and control samples more readily than in the corresponding cooked samples. This hypothesis allows the test to be one-tailed (Chapter 3).

In the triangle tests described in previous paragraphs, the experimental design stretched testers' acuity to the limit. Further trials using two test and one control sample could have improved the experimental design considerably, but sensory fatigue and probable lack of co-operation by participants would have been almost inevitable.

In the series of trials to determine optimal internal temperature, there were no complaints from tasters about the number of presentations. However, even if palate cleaners are provided, sensory acuity falls off when, in the interests of experimental design, too many samples are presented on any one occasion. With meats, the quantity available with identical characteristics is an additional constraint. Hence some compromise is inevitable. Therefore an investigation of the minimal number of replicates required in the tasting tests was made.

Results in paired comparisons and triangle tests are binomially distributed. Thus the binomial expansion may be used to determine the number of replicates required in each series of trials. As the number of replicates increases, the probability of identifying the odd sample correctly by chance alone on all occasions decreases. The same applies, but to a lesser extent, in paired comparisons.

In/

In paired comparisons, the probability of either sample being preferred on alloccasions if there are four replicates is <0.0625; where $p = q = 0.5$. To achieve this result, each participant must taste eight samples. Since triangle tests on the same samples were required at each session this probability was considered acceptable although by no means ideal.

For triangle tests used to detect differences in flavour, if four replicates are used, each taster would be required to taste an additional 18 samples making 26 in all. Whilst the probability of identifying the odd sample correctly by chance alone on all four occasions is <0.005 and on three of the four occasions <0.057, it seemed unrealistic to request that so many samples be tasted on a single occasion even if sufficient meat were available. Thus, in the early experiments it was decided to restrict triangle tests to two. Performance would subsequently be reviewed and, in designing the experiments, provision was made to carry out four triangle tests. In the event, after the first four sessions of the grass/rape trials on gigots, in the two sessions allocated to the assessment of loins there were four replicated triangle tests.

In the triangle tests used to identify possible differences in aroma of raw and cooked samples, there were three trials. The probability of identifying the odd sample correctly on all three occasions is <0.037. This was considered to be a satisfactory experimental design.

Probabilities were calculated for triangle tests on the assumption that a trial is one set of either two, three or four triangle tests. Each trial required three judgements. Each judgement has a one in three chance of being correct. The full

design/

design of the experiments, with either two or four triangle tests is indicated in Table 6.2. Although tasters' recording forms followed this design to avoid errors in punching in results for analysis the actual presentations to tasters were randomised in flavour trials between sessions and in the aroma trials it was indicated that judgements need not be made in any particular order (Plates 5 & 6 and Table 6.3 of the Appendix).

Table 6.2 Sensory Appraisal Tests - Grass Versus Rape : Experimental Design

Paired Comparisons

Pair	1	2	3	4
Sample A Codes	682	145	251	956
Sample B "	561	363	324	178

Triangle Tests - Tasting

Group	ABA	BAB	ABA	BAB
Sample A Codes	201		791	
Sample A "	855	861	144	119
Sample B "	725	634	432	818
Sample B "		412		556

Triangle Tests - Aroma

Group	1	2	3	4	5	6
Sample A Codes	945	954	426	237	516	532
Sample A "	382	195	779	488	736	213
Sample B "	538	221	141	275	189	918
R/C	C	R	R	C	R	C
	BAA	AAB	ABA	AAB	BAA	ABA

Abbreviations are those used in Table 6.1.

Following the assessments, the data was analysed with computer support. Specimen printouts are presented in the Appendix (Tables 6.1 and 6.2).

For/

For both paired comparisons and triangle tests, on the null hypothesis, results are binomially distributed. In the first analyses, values for χ^2 were calculated for both tests. Results which were of statistical significance were recorded for each experiment with df (K - 1). The binomial distribution allowed calculation of the expected frequencies in each category based on the null hypothesis. Actual and expected frequencies were compared. The Kolmogorov-Smirnov test (which compares cumulative frequencies of binomially distributed data) was then used to determine if the critical values of D enabled the null hypothesis to be accepted or rejected.

The results in each experiment were then grouped. Until recently this was generally accepted practice but as Harries (1982) indicates this procedure is open to criticism in that using n tasters once is that the same as using one taster n times. Only when the probability of a correct answer is the same for all subjects and independent from trial to trial is this practice valid. Where differences between meat samples is slight, a panel will comprise discriminators and non-discriminators. (This problem is discussed further in Chapter 3 and 13). Results from this grouped data are presented for interest although their limitations are recognised.

As in the experiments to determine optimal internal temperature, individuals' performance in each test was examined. Individuals who identified the odd sample correctly in at least 50% of the presentations, i.e. those who demonstrated aptitude in identifying differences were studied. There is little published work in this area, (Land D G, personal communication), but it could be argued that the preference judgements of such individuals are thus made on a more rational basis.

A/

A study was also made of those who preferred A or B on either three or four occasions, i.e. those who were consistent in their judgement. It has been suggested (Amerine et al 1965) that the inexperienced may perform better in paired comparisons than in triangle tests. It is recognised that preference is highly subjective. A sample considered by one subject to have inferior flavour may be preferred by another. Consistent preference is however of interest particularly as in the design and execution of the experiments efforts were made to minimise positional bias.

As indicated earlier in this chapter, Wilcoxon's Matched-Pairs Ranked-Sign Test was used to determine if differences between test and controls was more readily detected in raw than in cooked samples.

Files were created for the six sessions. The first four sessions compared gigots and the remaining two loins. Values of χ^2 are indicated in Table 6.3 (df 4). Where the calculated value in the paired comparisons exceeds 9.49, 13.28 and 18.46 the probabilities of achieving such results by chance alone are <0.05, 0.01 and 0.001 respectively. These values also apply to triangle tests on loin samples. Corresponding values of Chi square for triangle tests on gigots are 5.99, 9.21 and 13.82 (2df) with probability values of <0.05, <0.01 and <0.001 respectively.

Table 6.3 Values of χ^2 in Paired Comparison (Preference) and Triangle Tests : Grass versus Rape Flavour

File TA*	Cut	N	Paired Comparisons	Triangle Tests
A	G	17	16.33 **	0.07
B	G	18	6.44	3.63
C	G	15	7.04	1.20
E	G	17	2.45	0.60
D	L	13	2.59	3.81
F	L	13	5.77	4.98

The only value of statistical significance ($p<0.01$) is underscored.
The values of χ^2 for the aroma of raw and cooked samples are shown in Table 6.4 below.

Table 6.4 Values of χ^2 in Triangle Tests - Aroma of Grass versus Rape Samples Raw and Cooked

File TA*	Cut	N	Triangle Tests (Raw)	Triangle Tests (Cooked)
A	G	17	1.33	3.05
B	G	18	1.44	4.31
C	G	15	0.53	12.90 **
E	G	17	3.65	5.04 ($p<0.10$)
D	L	13	1.28	0.93
F	L	13	1.80	2.92

When the calculated values of χ^2 are 7.82, 11.34 and 16.27 (3df), they correspond to probabilities of <0.05, <0.01 and <0.001 respectively. The only trial where a statistically significant difference in aroma of grass and rape fed samples ($p<0.01$) is underscored.

The gigot sample results in files A and C and the loin sample results (which are in files D and F) relate to trials carried out at ESCA. Two of the gigot trials - files B and E - were carried out at QMC.

The results of these trials suggest that if differences in flavour or aroma exist between samples they are so small that they cannot, with one exception, be detected. Only one trial indicated a statistically significant flavour preference. This preference was for a gigot roast from a grass fed wether.

To confirm these findings, critical values of D which allow the null hypothesis to be accepted or rejected were calculated using the Kolmogorov-Smirnov Test. The expected frequency of responses in each category was calculated according to the binomial distribution. This test compares the cumulative expected and observed frequencies for each category. For paired comparisons there are five categories. There/

There are also five categories for the triangle tests on the flavour of loin samples. For triangle tests on the flavour of gigots and on the aroma of raw and cooked samples there are three or four categories respectively. The calculated values of D are indicated in Table 6.5.

Table 6.5 Grass versus Rape - Kolmogorov-Smirnov Test : Critical Values of D

File TA*	Cut	N	PCP	Triangle(F)	Triangle(AR)	Triangle (AC)
A	G	17	0.229	0.024	0.082	0.118
B	G	18	0.189	0.167	0.094	0.094
C	G	15	0.220	0.093	0.040	0.160
E	G	17	0.141	0.082	0.200	0.153
D	L	13	0.085	0.253	0.131	0.085
F	L	13	0.100	0.123	0.085	0.146

where PCP = Paired Comparisons (Preference)
F = Flavour
AR = Aroma raw
AC = Aroma cooked

From Table 6.5 it will be noted that none of the calculated values of D are large enough to allow the null hypothesis to be rejected. The observed results do not differ significantly from the expected results. Both are binomially distributed. The results of single trial of statistical significance indicated in Tables 6.3 and 6.4 have not been confirmed. These preliminary analyses indicate that the conclusions following these two tables are likely to be correct.

The total flavour preferences for sample A, the control, the calculated value of the z statistic and the probabilities for the grouped data are presented in Table 6.6.

Table 6.6/

Table 6.6 Paired Comparison (Preference) Tests: Grass versus Rape

File TA*	Cut	N	A preferred	Total	z	Probability<*
A	G	17	45	68	2.545	0.011
B	G	18	32	72	0.825	0.412
C	G	15	22	60	1.936	0.052
E	G	17	30	68	0.849	0.395
D	L	13	24	52	0.416	0.674
F	L	13	27	52	0.139	0.897
Gigots		67	129	268	0.55	0.582
Loins		26	51	104	0.098	0.920
Totals		93	180	372	0.674	0.502

Values which are of statistical significance are underscored.

The probability in Trial C is somewhat marginal.

The result for A, using the grouped data, confirms the χ^2 result for this trial. No clear pattern of preference for either sample is demonstrated by gigots or loins since in Trial A, A was the preferred sample whilst in Trial C, B was preferred.

Similar data for the triangle tests (flavour) is given in Table 6.7.

Table 6.7 Triangle Tests Flavour : Grass versus Rape

File TA*	Cut	N	Correct	Total	z	Probability<
A	G	17	11	34	-0.303	0.382
B	G	18	14	36	0.530	0.298
C	G	15	12	30	0.581	0.281
E	G	17	10	34	-0.667	0.251
D	L	13	12	52	-1.716	0.043
F	L	13	20	52	0.637	0.261
Gigots			47	134	0.366	0.356
Loins			32	104	-0.659	0.254
Totals			79	238	-0.146	0.440

The only value which appears to be of statistical significance is underscored/

*Two tailed test. Value in probability tables was therefore doubled to obtain these results.

underscored. In fact it would be incorrect to deduce that the odd sample of the three has been identified correctly on sufficient occasions for this result to be of statistical significance. The proportion of correct identifications by chance alone would have been at least 17 whereas the number of correct identifications was only 12 (23%). A similar result is obtained from the data of files TAA and TAE (32 and 29%). Such a peculiar pattern with a preponderance of incorrect responses in three of the six trials violates the assumption on which the triangle testing procedure is based. Neyman (1950), Bennett and Franklin (1954) and Byer (1964) discuss experimental findings of this type. It is suggested that this warns of some confusing bias in the experiment. If such a bias was in operation during the trials it was not possible to determine its cause. It should however be noted that these three trials were carried out at ESCA. Although volunteers had been carefully briefed they lacked experience in using sensory appraisal techniques in comparison with many tasters at QMC. This lack of experience may perhaps have been the interfering factor in the trials.

Performance of individual tasters in the paired comparison and preference tests was studied. Gigots and loins are considered separately since there were only two triangle tests carried out on gigot samples. This makes it more difficult to assess aptitude than in trials which followed when there were four tests in each trial. Subjects are generally considered to demonstrate aptitude for the test if the odd sample is correctly identified in at least 50% of presentations. In preference tests, subjects are judged to show consistency if either sample A or sample B is preferred on 3 or 4 occasions. Table 6.8 indicates consistent performance of subjects for both gigot and loin samples.

Table 6.8/

Table 6.8 Consistent Preferences for Grass and Rape Samples

File	TA*	Cut	N	4A	3A	3B	4B	A	B	Total
	A	G	17	5	4	3	0	9	3	12
(QMC)	B	G	18	2	4	6	3	6	9	15
	C	G	15	1	1	5	3	2	8	10
(QMC)	E	G	17	1	2	4	2	3	6	9
	D	G	13	0	3	5	0	3	5	8
	F	G	13	2	2	1	2	4	3	7
Totals								27	34	61

In addition, there were 32 occasions on which A and B were equally preferred. The final column of Table indicates that 61 subjects were consistent in their preference. Performance in triangle tests of these subjects was studied. Table 6.9 indicates the performance of these 61 subjects in triangle tests.

Table 6.9 Performance of Consistent Subjects in Triangle Tests: Grass versus Rape Trials

Occasions Correct	Gigots	%	Loins	%
0	20	43.5	4	27
1	20	43.5	7	47
2	6	13.0	1	7
3	NA		3	20
4	NA		0	0
Totals	46	49.5	15	27

where NA = not applicable

Thus of these 61 subjects only 49.5 and 27 percent showed aptitude for triangle tasting procedures. The number of subjects who prefer B to A (Table 6.8) must therefore be viewed with considerable caution. Preference is highly subjective and it is desirable that it should be established on a rational basis. This series of results does not indicate that this is so.

The results can be viewed differently if preferences of only selected/

selected tasters are considered, i.e. those who identified correctly in triangle tests on at least half the presentations. This information is set out in Table 6.10. For reasons already explained, gigots and loins are considered separately.

Table 6.10 Preferences of Selected Tasters for Grass or Rape Fed Samples

TA*	N	Triangle Tests Correct 1	2	Total	4B	3B	3A	4A	Subjects Preferring A	B
A	17	7	2	9	0	2	1	5	6	2
B	18	12	1	13	3	4	3	1	4	7
C	15	6	3	9	1	2	0	1	1	3
E	17	6	2	8	0	3	0	0	0	3
		Correct 2	3	4						
D	13	1	1	0	2	0	1	0	0	1
F	13	2	3	0	5	0	1	1	2	1
Totals	59				37	4	13	5	8	13	17

The total number of selected tasters is indicated in the centre of Table 6.10. Subjects who preferred either A or B do not correspond to these values. This arises because some of these subjects preferred A and B on an equal number of occasions. It is of interest to note that the slight preference for the flavour of B is maintained although the data has been analysed differently.

Differences in the aroma of raw and cooked samples were assessed by triangle tests. The design of the experiment is indicated in Table 6.1. The values of χ^2 and the probabilities are indicated in Tables 6.11 and 6.12

Table 6.11/

Table 6.11 Triangle Tests - Aroma of Raw Samples

File TA*	Cut	N	χ^2	Correct	Total	z	Probability‹
A	G	17	1.33	16	51	-0.466	0.326
B	G	18	1.44	16*	54	-0.722	0.236
C	G	15	0.53	15	45	-1.58	0.436
E	G	17	3.65	12*	51	-1.634	0.052
D	L	13	1.28	16	39	0.849	0.198
F	L	13	1.80	12	39	-0.510	0.305
Totals G				59*	201	0.360	0.359
Totals L				28	78	-1.272	0.102
Totals G + L				87	278	-0.826	0.203

*Correct responses for these groups were only 29.6%, 23.5% and 29.4% respectively.

Thus there are no detectable differences in the aroma of samples from grass fed or rape fed lambs. The only result, File TAE, which appears to be of statistical significance is not valid because in only 23.5 per cent of presentations was the correct identification made. As indicated previously, the fundamental assumption of the triangle test is violated.

Table 6.12 Triangle Tests - Aroma of Cooked Samples

File TA*	Cut	N	χ^2	Correct	Total	z	Probability‹
A	G	17	3.05	17	51	-0.149	0.480
B	G	18	4.31	17	54	-0.433	0.334
C	G	15	12.90[d]	21	45	1.739	0.041[d]
E	G	17	5.04	18	51	-0.144	0.444
D	L	13	0.93	16	39	-0.849	0.198
F	L	13	2.92	12	39	0.849	0.198
Totals G				27	78	0.823	0.206
Totals L				73	201	0.120	0.452
Totals G + L				100	278	0.826	0.203

where d = $p<0.05$

In the cooked samples, apart from one trial, no statistically significant differences in aroma were demonstrated. In one trial (File/

(File TAC), subjects' response patterns were 2, 8, 2, 3 for zero, one, two or three correct identifications respectively in the three presentations. These participants (ESCA) were carrying out their first series of trials and were likely to have been well motivated. This may have encouraged them to identify differences where none existed. This apparent difference could have arisen as a result of changes induced by the cooking process since only $33^{1}/3$ per cent of correct identifications were made in the raw samples. It thus seems likely that detectable differences in the aroma of cooked samples between grass and rape fed animals are absent.

As indicated earlier in this chapter, Wilcoxon's Matched-pairs Signed-Ranks Test was used to test the hypothesis that differences would be more readily detected in raw than cooked samples when the triangle test is used. The scores for each taster, 0, 1, 2 or 3 were compared for raw and cooked samples. Results of this analysis are given in Table 6.13. It will be noted that none of the observed values of T is equal to or less than critical values set out in Table J (Haber and Runyon 1977).

Table 6.13 Wilcoxon Matched-Pairs Signed-Rank Tests: Grass versus Rape I

File TA*	N	DS	T	Plus	Minus	Minus<Plus	Result
A	17	14	51.5	51.5	53.5	≡	NS
B	18	13	44.5	44.5	46.5	≡	NS
C	15	10	16.5	16.5	38.5	Yes	NS
D	13	9	13.5	31.5	13.5	No	NS
E	17	12	25.5	25.5	52.5	Yes	NS
F	13	7	7	7	21	Yes	NS

where N = number of tasters
DS = total number of observations having a different sign
T = smaller of like signed ranks

Thus/

Thus the hypothesis that detection of the odd sample in triangle tests would occur with greater frequency in raw than in cooked samples must be rejected. In Table 6.12, it was indicated that detectable differences in the aroma of cooked samples may be present in one trial (File TAC). In the present test the performance of each individual in the two tests is assessed separately. Hence it does not follow that this test will be in agreement with previous tests.

3. Discussion and Conclusions

Only one statistically significant flavour preference was established by Paired Comparison (Preference) tests. This preference was not associated with flavour difference as assessed by triangle tests. Similarly a statistically significant flavour difference was not associated with preference for either sample. There were thus no demonstrable differences between flavour or aroma from this batch of grass and rape fed sheep reared on ESCA farms. Lambs were selected as far as possible for uniformity. Feeding regimes were recorded and controlled. Slaughter, frozen storage, thawing and cooking procedures were standardised throughout. The latter followed conventional techniques as practised in the United Kingdom for these cuts of meat. Thus in these trials the findings of Park et al (1972b) were not confirmed.

Because of the experience of these workers in recording flavour defects in lamb samples it was considered that the Grass versus Rape trials should be repeated. That these findings could not be confirmed could be caused by any or a combination of any of the following:

1. Tasters lacked the necessary experience, expertise and aptitude for detecting differences by the use of triangle tests.
2. That there was little or no detectable difference between grass and rape fed lambs in these trials.

3./

3. It may not be possible to transfer findings of workers outside the United Kingdom in that:

   (i) Breeds of sheep differ

   (ii) Husbandry/Slaughter practices differ

   (iii) Forage crops - although of the same variety and cultivar - may differ according to locality as a result of variation in climatic and soil conditions

   (iv) Different cooking procedures were used

   (v) Subjects were not requested - as in these trials - to make only a hedonic scale rating. Verbal 'clues' were also supplied in other trials (Park et al 1972a)

   (vi) Subsequent qualifications by these workers. (Chapter 9. Developments in Meat Science, Vol.1. Ed. R A Lawrie 1980 Academic Press)

   (vii) MRI Annual Report of 1974 (anon) in which differences between grass and rape fed animals could not be confirmed).

It is thus concluded that either no detectable differences existed between samples or that tasters were not performing adequately. Although tasters were inexperienced, they appeared to be well motivated and eager to participate in the trials. Two experiments were therefore devised to follow these trials. In the first, two highly trained and experienced tasters would carry out a Flavour Profile Analysis on gigots and loins from two rape fed samples and one grass fed sample. Should variation between the two rape fed samples equate with that of the grass fed sample, it would seem unlikely that differences would be detected by the use of triangle and paired comparison tests with inexperienced tasters. The results of these testing procedures are considered in Chapter 8 (These results suggest that it was unlikely that differences could/

could be detected between the two feeding regimes and that the results of the present series of trials could only have been expected). Joints from the same grass and rape fed samples were used.

In addition, tasters' performance in testing procedures would be assessed by the use of two very different meats. As well as assessing their performance, tasters would be likely to be encouraged by greater success in detecting differences. The results of this series of experiments is presented in Chapter 7.

Thus in repeating Grass versus Rape Trials combined with assessing tasters' acuity it was considered that valid estimations of any differences in eating quality of grass and rape fed samples could be confirmed or refuted. The only reservation concerning such experiments is that the number of carcasses sampled was inevitably small.

In these Grass versus Rape Trials, one grass fed gigot was compared with one rape fed gigot. To provide enough meat for tasters, it was necessary to pool samples from two loins. This was justified on the assumption that flavour differences between feeding regimes would be greater than differences within the same regime. In trials described in Chapter 9, because of the greater numbers of tasters expected, two gigots and four loins from each feeding regime were pooled.

CHAPTER 7

Summary - Assessing Tasters' Performance

Following the same procedures as in the first series of grass versus rape trials, May slaughtered barley/protein supplement fed females reared indoors were compared with turnip fed wethers slaughtered the previous February. Statistically significant differences were demonstrated between samples. The majority of tasters are consistent in their preference with B (the turnip fed wethers) being the preferred sample. Statistically significant flavour differences were demonstrated by triangle tests. Differences in aroma were also established with such differences being more readily detected in raw samples. It can be concluded that such tests can be used to demonstrated differences between samples and as a basis for assessing preference.

Establishing Tasters' Discrimination and Consistency of Judgements

1. Whilst it had already been decided to repeat the grass versus rape trials, it was considered important to check tasters' discriminatory ability and their consistency in replicated trials. Thus a comparison of samples likely to show considerable differences was made using the same experimental design and procedures as in the grass versus rape trials. 'Spring' lambs from females fed barley with a protein supplements (A samples) slaughtered in May were compared with joints from much older turnip fed wethers slaughtered the previous February (B samples). The A samples were Suffolk x Dorset/Finn females, born at the beginning of January and housed throughout their lives. They were weaned in March and from then until slaughter on 4th May 1978 were fed on a whole barley/protein pellet mix and a small supplement of hay. They/

They were thus four months of age at the time of slaughter.

The B (turnip fed) samples were Blackface wethers born in April/ May 1977 which had been kept on the hill until August. After weaning, they were grazed on lowland grass fields until 5th December when yellow turnips were used as fodder. Whole barley was introduced and gradually increased so that, at slaughter they were eating 350g/day. They were slaughtered on 14th February 1978 at approximately ten months of age.

Thus feeding regime was not the only variable. The younger animals were females of a different breed which had remained indoors throughout whereas the older animals were Blackface wethers reared outdoors on turnips and whole barley. Samples could thus be expected to be very different. Trials were started in June 1978. However, that sire or breed and rearing practices exert little effect on eating quality of lamb is indicated by Dransfield et al (1979) and Rhodes (1971).

The appearance of the raw meats - both the colour of the muscle and of the fat - showed marked differences. The meats from the younger lambs was lighter in colour, less fatty and juicier in appearance. These observations are not however of importance as far as tasters are concerned in that their judgements are made on cooked samples presented uniformly as indicated in Chapter 4. After cooking, preparation of samples of uniform appearance was made difficult in that:

1. The age of the animal at slaughter affects the temperature of the colour change in meats. Hence the extent of colour change at constant internal temperatures was slightly less in joints from younger animals.
2. As a result of lower myoglobin concentration, meat from younger animals was lighter in colour.

Whilst/

Whilst such colour differences were unlikely to affect assessment of aroma in triangle tests, identification of the odd samples, despite random positioning, was relatively easy in the triangle tests for flavour. However, perhaps the experimenter was more aware of the problem than the tasters. Tasters were eager to carry out tests with care and accuracy. Because they were inexperienced, they were probably more likely to follow instructions exactly. They were requested to carry out paired comparison (preference) tests and triangle tests only on the basis of flavour. Some might have been tempted to use colour difference to assist them to identify the odd sample in the triangle tests and colour differences could affect also judgements in preference tests. Nevertheless in the ESCA/QMC student trials described, three of the 17 subjects performed poorly in the triangle tests but were consistent in their preference.

Earlier tests were carried out in daylight. It was at this stage considered a serious disadvantage that there was no access to a sensory appraisal room with coloured lighting facilities which might have masked differences in colour. More complete attention to flavour preferences and identification of the odd samples without distraction might have been given. In these earlier experiments, results indicated that tasters showed both discriminatory ability and consistency in judgements. These results could perhaps have been viewed with greater confidence at the time had specialist accommodation been used. Apart from the final ESCA trial, the series of tests was completed by a group using a purpose-built sensory appraisal room with the facility for individual tasters to be isolated - and hence unaware of others' responses - and coloured lighting to be used. Although some improvement was noted, coloured lighting could not completely disguise differences/

differences in depth of pigmentation. These results were obtained in November 1978.

In determining optimal internal temperature, joints were cooked to 70°C, 75°C, 80°C and 85°C. Differences in appearance which were inevitable could have influenced judgements on the hedonic rating scales. It was thus only in the grass versus rape trials (Chapter 6) that the appearance of samples had been uniform in all the studies so far reported.

2. Results of the Testing Procedure

The first of the trials was made in June 1978. Initially the grouped data was assessed for paired comparisons, for triangle tests of flavour and aroma of raw and cooked meats. This implies (Harries & Smith 1982) that the probability of a correct identification in the triangle test is the same for all subjects and is independent from trial to trial. Only when there is no discrimination is it true that using one taster n times is the same as using n tasters once. As the limits of discrimination are reached i.e. when the magnitude of difference between samples is small, differences in the acuity and discriminatory ability of panel members will become evident as well as for an individual panel member. In this series of trials it was presumed that the majority of tasters should demonstrate acuity. Results are presented in the tables which follow.

Table 7.1 Results of Gigot (LHS) Samples in Paired Comparison (Preference) and Triangle Tests - Grouped Data

Table 7.1 Results of Gigot (LHS) Samples in Paired Comparison (Preference) and Triangle Tests - Grouped Data

Trial	Paired Comparisons			Triangle Tests (Flavour)	
	A preferred	B preferred	N	Correct	N
1	15	45	60	29	30
2	25	54	79	25	40
3	24	40	64	18	32
4	15	17	32	8	16
Totals	79	156	235	80	118
%	33.6	66.4		67.8	
χ^2 Values		24.47		61.53	
P		0		0	

Table 7.2 Results of Triangle Tests (Aroma) - Gigots

Trial	Raw		Cooked		
	Correct	N	Correct	N	
1	21	45	16	45	*
2	32	60	29	60	
3	23	48	17	48	
4	15	24	8	24	*
Totals	91	177	70	177	
%	51.4		39.55		
χ^2 values	25.23		22.2		
P	0		$p<0.046$		

*Note these results are approximately those which would have been achieved by chance alone, i.e. 1 in 3 of the odd samples would be expected to be identified.

Table 7.3 Results of Loins (LHS) Samples in Paired Comparison (Preference) and Triangle Tests

Trial	Paired Comparisons			Triangle Tests (Flavour)	
	A preferred	B preferred	N	Correct	N
5	7	17	24	16	24
6	12	12	24	22	24
Totals	19	29	48	38	48
%	39.6	60.4		79.2	
χ^2 values		1.69		43.33	
		$p<0.194$ NS		$p<0.001$	

Table 7.4 Results of Triangle Tests (Aroma) Loins

Trial	Raw		Cooked	
	Correct	N	Correct	N
5	5	18 *	9	18
6	8	18	10	18
Totals	13	36	19	36
%	36.1		52.8	
χ^2 values	0.781		5.281	
	$p<0.468$		$p<0.011$	

*Despite this small disparity, this number of correct identifications is less than $33^1/3\%$. There were however only six participants.

The results of these combined totals should be compared with those of Grass versus Rape 1 where no statistically significant differences were demonstrated.

When results of the individual trials are separated, the findings are indicated below.

Table 7.5/

Table 7.5 Results of Gigot (LHS) Samples in Paired Comparison (Preference) and Triangle Tests for Each Trial

Trial	Paired Comparisons					Triangle Tests (Flavour)			
	A preferred	B preferred	N	χ^2	Probability<	Correct	N	χ^2	Probability<
1	15	45	60	42.33	0	29	30	102.75	0
2	25	54	79	17.20	0.002	25	40	18.92	0
3	24	40	64	5.67	0.602	18	32	11.56	0.005
4	15	17	32	5.83	0.857	8	16	3.53	0.125

Table 7.6 Results of Triangle Tests (Aroma) - Gigots

Trial	Raw				Cooked			
	Correct	N	χ^2	Probability<	Correct	N	χ^2	Probability<
1	21	45	5.92	0	16	45	3.60	0.830
2	32	60	12.63	0.001	29	60	6.83	0.008
3	23	48	9.17	0.023	17	48	5.52	0.452
4	15	24	15.20	0.002	8	24	2.27	0.409

Table 7.7 Results of Loin (LHS) Samples in Paired Comparison (Preference) and Triangle Tests for Each Trial

Trial	Paired Comparisons					Triangle Tests			
	A preferred	B preferred	N	χ^2	Probability<	Correct	N	χ^2	Probability<
5	7	17	24	8.44	0.066	16	24	23.67	0.001
6	12	12	24	2.44	0.841	22	24	216.75	0

Table 7.8 Results of Triangle Tests (Aroma) - Loins

Trial	Raw				Cooked			
	Correct	N	χ^2	Probability <	Correct	N	χ^2	Probability <
5	5	18	3.94	0.225	9	18	6.94	0.105
6	8	18	15.75	0.040	10	18	6.19	0.008

The actual probabilities were calculated for each test using the z statistic. It should however be noted that the values of χ^2 noted took into account the frequency with which subjects preferred a particular sample or identified the odd sample correctly in triangle tests. Results for trials 5 and 6 in particular where there were only six participants (although in other laboratories panel sizes of this order are considered acceptable and tasters were, by that stage, becoming experienced), should thus be viewed with caution.

For interest, combined figures for gigots and loins are shown in Table 7.9. Results which also include the results of the ESCA/QMC experimental session are also shown. Since figures for preferences and correct identification are already tabulated for individual trials, for grouped data on gigots and loins and for the experiment mentioned in the previous paragraph they have been omitted in the table which follows.

These results were achieved by using 60 tasters. Of the 60, 18 took part on two occasions, two on three and two on four occasions. One subject in the latter category performed with remarkable consistency in the four sessions, particuarly in triangle tests. On only one occasion was A preferred by this subject. The other performed well in triangle tests but was inconsistent in paired comparisons. It is a matter of regret that further studies of this type were not possible.

Table 7.9 Probability Values in Paired Comparison and Triangle Tests

	Paired Comparisons	Triangles	Flavour	Aroma(R)	Aroma(C)
Gigots	0	0		0	0.047
Loins *	0.194	0		0.218	0.011
Gigots & Loins	0	0		0	0.006
All samples	0	0		0	0

where R = raw C = cooked samples

*See previous comment on this data following Table 7.8. As indicated previously, the actual probabilities were obtained by calculating the z statistic.

When the data is grouped in this way, combining the results of gigots and loins suggests that the probability of achieving them by chance is virtually nil in relation to the paired comparison preference tests. Clearly, subjects appeared to find little difficulty in identifying the odd sample in the triangle tests on the basis of flavour and showed the ability to do so. The proportion of correct identifications is extremely high.

It will be noted from the specimen print-out (Appendix Table 6.2) that the actual and expected frequency of responses is available. The Kolgomorov-Smirnov Test was used to determine if the actual scores were binomially distributed. The critical values for D are indicated in Table 7.10. Computer files TAG to TAL correspond to trials 1-6 described earlier in this chapter. The corresponding figures for the ESCA/QMC trials yet to be described are included for purposes of comparison.

It will be noted that in only two of the paired comparison (preference) tests does the maximum deviation (D) indicate that results are not distributed binomially.

Table 7.10 Acuity Tests: Kolmogorov-Smirnov Test - Critical values of D for TA*

File TA*	N	PC(P)	ΔF	ΔAR	ΔAC
G	15	.407	.827	.273	.160
		**	**	NS	NS
H	20	.340	.295	.340	.240
		*	*	*	NS
I	16	.188	.263	.300	.088
		NS	NS	NS	NS
J	8	.188	.325	.500	.100
		NS	NS	*	NS
K	6	.367	.567	.250	.417
		NS	*	NS	NS
L	6	.117	.900	.300	.417
		NS	**	NS	NS
T3Q	17	.123	.594	-	-
			**		

Where/

Where PC(P) = paired comparisons (preference) and

F, AR and AC are triangle tests for flavour, aroma raw and aroma cooked of samples respectively

* = $p<0.05$ ** $p<0.01$

NS = not significant

In four of the seven triangle tests for flavour results are not binomially distributed as are two of the six triangle tests for the aroma of the raw samples. The first two trials were carried out at QMC where participants had more experience of the testing procedures which could have influenced results. It was necessary to use critical values of D for a two-tailed test since values for one-tailed tests are neither available or appropriate to this statistic.

The Wilcoxon Match-Pairs Signed-Ranks Test was carried out on the results of the triangle tests of the raw and cooked aroma of samples. Following the previous experimental studies with pork, it was hypothesised that identification of the odd sample of the three would be more easily accomplished in raw samples. The information on these comparisons is presented in Table 7.11.

Table 7.11/

Table 7.11 Wilcoxon Matched-Pairs Signed-Ranks Tests

File TA	N	DS	T	+	-	-<+	Probability
G	15	11	20.5	45.5	20.5	No	NS
H	20	15	47.5	72.5	47.5	No	NS
I	16	10	14.5	40.5	14.5	No	NS
J	8	6	5	15	5	Yes	NS
K	6	5	2.5	2.5	12.5	Yes	NS
L	6	4	3	3	7	Yes	NS

Where N = number of participants

DS = total number of values with a different sign

T = smaller of the sums of like signed ranks

Thus for all gigot samples there was no difference in performance when raw or cooked samples were assessed. Tasters showed no greater discriminatory power in triangle tests when the aroma of samples was assessed before cooking. The experimenter's hypothesis was thus rejected.

Since preference for B is demonstrated, results were tabulated in Table 7.12. In Table 7.13 which follows preference for B is linked to subjects' performance in triangle tests (flavour).

Table 7.12 Subjects' Preference for Sample B

File TA*	4B	3B	2B	1B	Nil B	Total
G	7	3	3	2	0	15
H	5	8	3	4	0	20
I	3	5	5	3	0	16
J	2	1	2	2	1	8
K	2	2	1	1	0	6
L	1	1	2	1	1	6

Subject/

Subjects preferred B on 4, 3, 2, 1 or zero occasions on each of the 4 presentations. Column 2B is ignored in that preference appears to be random. Only 15 participants preferred A on either three or four occasions whereas 40 preferred B on three or four occasions. The performance of the 40 who preferred B in the triangle tests is indicated in Table 7.13 below.

Table 7.13 Linking Subjects' Preference with Performance in Triangle Tests TA*

	2	1	Nil	2	1	Nil
4B	7	0	0	1	3	1
3B	2	1	0	3	3	2

	G			H		
4B	2	1	0	1	1	0
3B	2	1	2	0	1	0

	I					J				
	4	3	2	1	0	4	3	2	1	0
4B	0	0	1	1	0	1	0			
3B	0	2	0	0	0	1	0			

	K	L

It will be noted that for loin samples the number of triangle test presentations was increased to four. If performance in triangle tests is good, this suggests that preference may have a rational basis. If it is considered that some aptitude is demonstrated by correct judgement on at least 50 per cent of occasions the majority of subjects who are consistent in their preference also perform well in triangle tests.

Table 7.14/

Table 7.14 Summarising the Results of Table 7.13

4B/3B	%	2Δ/1Δ or 4Δ/3Δ/2Δ	%	N	File TA*
10	67	10	67	15	G
13	65	10	50	20	H
8	50	6	38	16	I
3	38	3	38	8	J
4	67	4	67	6	K
2	33	2	33	6	L

In addition to the results presented in Table 7.14 it should be recognised that 12 subjects consistently preferred A. These 12 also demonstrated aptitude for the triangle testing procedures. They represented a very high proportion of 15 subjects who preferred A.

In conclusion, it may be stated that the majority of subjects showed some aptitude for triangle tests and were also consistent in their preferences for either B or A. From Table 7.12, the small number of participants who selected A and B on an equal number of occasions is demonstrated. They account for 16 of the responses in comparison with 40 consistent preferences for B and 14 consistent preferences for A. Those who are consistent in their response clearly exceed those who demonstrate no preference for either sample.

3. The observations in Table 7.15 were derived from the computer printout from file T3Q. This file gives the results of the experiments carried out by 17 ESCA/QMC students. It forms the basis of the discussion of the experimental results which follows.

Table 7.15/

Table 7.15 The Computer File T3Q - ESCA/QMC Students 14.11.78

	Paired Comparisons N = 4	Triangle Tests N = 4
201	4	3
202	3	4
203	3	3
204	1	4
205	4	1
206	2	4
207	3	2
208	2	4
345	1	1
346	2	4
347	0	4
348	0	3
349	1	1
350	0	3
351	2	2
352	4	4
353	2	3
χ^2 (4df)	8.10 ($p<0.10$)	233.49 ($p<0.00$)

The first column identifies the taster. The second column (paired comparison preference tests) indicates the number of occasions in the four presentations on which Sample B was preferred. It can be seen that responses thus fall into five categories. The third column indicates the number of correct responses in the triangle tests. These also fall into five categories. Also available (not set out here) is a table summarising the link between tasters' performance in paired comparison and triangle tests, for example, of the three tasters who consistently preferred Sample B i.e. on all four presentations, correct identifications of the odd sample in the four triangle tests were made on one, two, three and four presentations respectively. This table has greatly simplified the more detailed analyses of the data which follows.

In/

In November 1978, a group of 17 students from both ESCA and QMC took part in an additional series of tests. Four replicated paired comparison (preference) tests and four replicated triangle tests were used. Individual booths in the Sensory Appraisal Room were available for the tasting tests. Both orange and white tungsten filament lighting was used. Both groups of students had had some experience in this type of tasting procedure. The experimental design is indicated below:

Orange Lit Booths Sample Codes				White Lit Booths Sample Codes			
Paired Comparisons							
682	A	363	B	251	A	178	B
1st		2nd		3rd		4th	
561	B	145	A	324	B	956	A

Triangle Tests

1st	201	725	855	2nd	634	861	412
	A	B	A		B	A	B
3rd	791	432	144	4th	818	119	556
	A	B	A		B	A	B

Where A = Spring (May slaughtered lamb) B = February slaughtered lamb

The results of these tests are indicated in Table 7.16. Totals for paired comparisons and triangle tests under each lighting condition and presentation sequence are indicated.

Table 7.16 Results of Paired Comparison (Preference) Tests

Paired Comparisons			Triangle Tests		
Light Source	A	B	Light Source	C	I
Orange AB	9	8	Orange ABA	12	5 **
Orange BA	5	12	Orange ABA	11	6 **
White AB	10	7	White BAB	13	4 ***
White BA	10	7	White BAB	14	3 ***

Totals/

Totals Orange A 14 B 20 C 50 I 18 ***

White A 20 B 14

Where C = Correct identification

I = Incorrect identification

** = $p<0.01$

*** = $p<0.001$

NS = Not significant

It will be noted that there appears to be a reversal of preference in the paired comparisons when the total scores are considered. Similarly, in the triangle test, acuity appears to be greater in white than in orange light.

Examination of Table 7.17 which follows indicates that these grouped results are misleading. Preference is highly subjective. Three people consistently preferred Sample A irrespective of lighting conditions. Similarly four people consistently preferred B. It is of interest to note the reactions of the five subjects who preferred A and B on an equal number of occasions. Two followed the apparent trend in that B was preferred in orange light and A in white light. The remaining three were unaffected by lighting conditions always choosing with inconsistency.

Table 7.17 Performance of Five Subjects Preferring A and B on an Equal Number of Occasions

	Orange Lighting		White Lighting	
Subject 6	2B		2A	
" 8	1A	1B	1A	1B
" 10	2B		2A	
" 15	1A	1B	1A	1B
" 17	1A	1B	1A	1B

In addition there were three subjects who chose A on three of four occasions/

occasions and also three who chose B on three of four occasions. Like the first seven subjects discussed, their consistency was good. Thus apart from the five subjects considered in Table 7.17 the experiment demonstrated consistent preferences for either A or B by 10 of the 17 participants. Even of the five inconsistent subjects, two can be considered to be consistent in their preference although this preference varied with light source.

Performance of these 17 subjects in the triangle tests, with three exceptions, is good. (Aptitude for the testing procedure is considered to be demonstrated if the correct identification is made on at least 50% of occasions). Details of performance of each subject in these tests is given in Table 7.18 and in the Paired Comparisons in Table 7.19. In relation to triangle tests, 12 subjects identified the odd sample correctly on either three or four occasions irrespective of lighting conditions. Of the two subjects who identified the odd sample correctly on two of four occasions, one could identify the odd sample only in orange light and the other was unaffected by lighting conditions. Subjects 5, 9 and 13, who identified the odd sample only once, all achieved this single identification only under white light. Lighting conditions do not thus seem to exert much effect in the triangle tests.

Linking the paired comparisons and triangle tests for these three subjects, subject 5 consistently preferred B and both subjects 9 and 13 preferred A on three of four occasions. Their performance is very much in line with findings in the experiment "The Visual Impact of Foods" where subjects were more consistent in their preferences than in identifying the odd sample in triangle tests. To quote a student comment - "they know what they like but not why they like it!"

The series of trials of the barley fed May slaughtered lamb and the/

Table 7.18 Analysis of Performance - Triangle Tests

Subject	Sample Codes 725	861	432	119	$^{x}/_{4}$	Incorrect/Light Source
1	1	1	1	0	3	W
2	1	1	1	1	4	
3	0	1	1	1	3	0
4	1	1	1	1	4	
5	0	1	0	0	1	W
6	1	1	1	1	4	
7	1	0	1	0	2	2W
8	1	1	1	1	4	
9	0	0	0	1	1	W
10	1	0	1	1	3	W
11	1	1	1	1	4	
12	1	1	0	1	3	0
13	0	0	0	1	1	W
14	0	1	1	1	3	0
15	1	0	0	1	2	
16	1	1	1	1	4	
17	1	1	0	1	3	0
Light Source	0	W	0	W		
Totals						
Nos. correct	4	3	2	1	0	
	6	6	2	3	0	
Errors		0 4				
		W 2				

where:

Correct Responses 1
Incorrect Responses 0
Light sources 0 = Orange
W = White

Table 7.19 Analysis of Performance - Paired Comparison (Preference)

	Tests					
Subject	1st	2nd	3rd	4th	B preferred	Triangle Tests ×/4
1	B	B	B	B	4	3
2	B	B	A	B	3	4
3	B	B	B	A	3	3
4	A	B	A	A	1	4
5	B	B	B	B	4	1
6	B	B	A	A	2	4
7	A	B	B	B	3	2
8	A	B	A	B	2	4
9	B	A	A	A	1	1
10	B	B	A	A	2	3
11	A	A	A	A	0	4
12	A	A	A	A	0	3
13	A	A	B	A	1	1
14	A	A	A	A	0	3
15	A	B	B	A	2	2
16	B	B	B	B	4	4
17	A	B	A	B	2	3
Light Source	0	0	W	W		where 0 = orange W = white light
Nos. choosing B	4	3	2	1	0	
	3	3	5	3	3	
	1	1		1	1	4 Performance
	1	1			2	3 in Triangle
		1				2 Tests
	1			2		1
						0

the February slaughtered lamb were completed in January 1979 by two groups of ESCA BSc students of Agriculture. Experience since 1976 had shown students to have good discriminatory ability even in the far from ideal conditions in which the tests were carried out. In 1979, the students obtained statistically significant results in paired comparison (difference) tests on beef samples, in triangle tests on the aroma and flavour of young and mature pork and for juiciness in lamb samples cooked to internal temperatures of either 70°C or 80°C.

For this particular series of trials, students were first required to indicate preference for A or B on the basis of flavour. The same samples were then, unknown to the students, assessed on the basis of general acceptability. Presentation of samples followed the usual practice. Results of the tests are shown in Table

Table 7.20 Results of Paired Comparison Tests

Basis of Choice	N	A	B	% Preference B	χ^2	z	p
Flavour	56	19	37	66	5.2*	2.27	<0.012
Acceptability	56	22	34	61	2.2	1.46	<0.072

Where * = $p<0.05$

The probability of preference for the flavour of B by chance is very slight. Even when other characteristics are considered, as in general acceptability, the probability of achieving these results by chance is low. (In only 7% of trials would a different result have been obtained). In this series of trials, it was thus established that tasters preferred sample B and that their preference was more obvious when flavour alone formed the basis of preference.

Of the 56 participants, 33 (59%) consistently preferred either Sample A or Sample B in both tests. It is thus suggested that, on this occasion, the 33 showed greater aptitude for the tests. Their results/

results are shown in Table 7.21.

Table 7.21 Results of Tasters Consistent in Flavour and General Acceptability

N	A	B	% Preference B	χ^2	z	p
33	9	24	73	5.94	2.44*	<0.007

4. Discussion and Conclusions

The results of all the trials carried out seem to indicate that if sufficient difference between samples exists a relatively high proportion of subjects are able to identify odd samples in triangle tests and to make consistent preference judgements for either of the samples in paired comparison (preference) tests. Three very different groups of tasters were used for these trials. The ESCA/QMC students were less strong in their preference for Sample B. Preference is however highly subjective and in this group were tasters who either consistently preferred boar meat or conventional pork in 4 replicate studies. They may not therefore be quite as representative of the typical consumer as the other two groups.

It must however be admitted that the general trend for B to be the preferred sample was a matter of surprise to the experimenter. It was the time of year when the flavour of lamb is generally being described as "too strong" and the consumer is prepared to pay a higher price for "spring lamb". Yet the older lamb was clearly the preferred sample in this series of trials.

Referring again to the first series of grass versus rape trials, it becomes possible to feel more reassured that the fact that statistically significant differences were not demonstrated was not a result of poor performance by assessors. The latter have shown themselves able to carry out this type of testing procedure very effectively.

Rather/

Rather it must be that any differences, if present, were so slight as to be undetected in the conditions of these experiments. In addition, it may be that differences in cooking samples combined with differences in testing procedure and the analysis of data have shown that demonstrable differences do not exist between grass fed and rape fed animals reared in Scotland.

This series of trials has thus fulfilled its purpose in vindicating the tasters.

CHAPTER 8

Summary - Grass versus Rape: Second Series of Trials

Triangle tests indicated that there could have been slight differences in flavour between roasted gigots and loins from grass and rape fed lambs. The flavour of grass fed lambs was preferred. Results however indicate that it is unlikely that most consumers would detect differences between joints from the two feeding regimes despite the ability of second year BSc students of Agriculture to detect aroma differences in raw and cooked samples.

1. Grass versus Rape: Second Series of Trials

Despite greater precision of experimental design, no statistically significant differences between or preferences for gigots and loins from either grass or rape fed samples were established in the first series of trials. These results could either be caused by the absence of detectable differences between test and control samples or to lack of acuity of panel members, many of whom were still relatively inexperienced.

Since first series samples were still available it was decided that they should be tested by descriptive sensory techniques. Two of the author's colleagues have been trained in the flavour profile analysis technique at the Arthur D Little Institute (Inc), U.S.A. Both have had considerable experience and regularly undertake consultancies relating to product development, quality control and problems of taint in the food industry. Although this technique is normally used where detectable differences between samples have been demonstrated, they agreed to carry out analyses on roasted gigots and loins from grass and rape fed animals.

Samples/

Samples of meat were prepared as for earlier panel trials for each taster. For analysis of aroma, meat samples were placed in 250 ml tall form Pyrex beakers covered with watchglasses. The Sensory Appraisal Room was used.

From their report, it is evident that there were few characteristics in common in the samples from the two rape fed animals and little to differentiate them from the grass fed samples. On each occasion, they considered that they were testing three different samples rather than two similar/identical and one odd sample. Since the joints were well within the limits of recommended frozen storage times for lamb, it seems unlikely that the additional storage would have induced changes sufficient to mask differences between samples. Loins were assessed on the first occasion and gigots on the second. Their report is presented in Table 8.1 of the Appendix.

In view of the nature of the flavour profile analysis technique it is thus hardly a matter of surprise that less experienced tasters were unable to identify the odd sample correctly or to make consistent preference judgements. Thus it can be reasonable to conclude that, despite the findings of other workers (Park et al 1972), there was no detectable difference between rape and grass fed samples on this occasion.

2. In Chapter 7, tasters' performance has been shown to be vindicated. When differences exist between samples these are detected by triangle tests. The majority of subjects are consistent in their preferences. These findings, combined with the flavour profile analyses, enabled the second series of trials to be carried out with greater confidence in the methodology and in the aptitude of tasters for the testing procedures.

The/

The aroma tests of the first series had proved most unpopular. They were considered by participants to be extremely difficult, disagreeable and off-putting to an extent that they hesitated about continuing with the series of trials. Results were in no way as clear cut as those which had been achieved in the de-oiled herring silage trials. A decision was reached that they should be omitted from the second series of trials. Otherwise, triangle and four paired comparison (preference) tests were used as in the first series (roasted loins). There were four replicates in the triangle tests. For the second series of trials, both grass fed and rape fed animals were Scottish Blackface wethers. All the animals were weaned in August and grazed on grass aftermaths until 5th October 1978. The rape (B) group were moved on to their crop whilst the remainder (A group) stayed on grass. No supplementary feeding was offered to either group. The slaughter date of both groups was 20th November. The slaughter weights of each group are indicated in Table 8.1.

Table 8.1 Slaughter Weights of Grass and Rape Fed Blackface Wethers used in Second Series of Trials

Grass		Rape	
Animal No.	Slaughter Weight (kg)	Animal No.	Slaughter Weight (kg)
555	40	W153	39.5
1310	39	W138	39
787	40	0382	38
569	38	1468	39*
90	37		39
975	40.5		

With the exception of the asterisked sample, where the condition score was 3, the others were rated as $2\frac{1}{2}$.

These/

These animals were sold to FMC. Hence individual carcass weights are not available. Assuming 46-47% live weight conversion to carcass (J FitzSimons) for lambs of this age estimated carcass weights are indicated in Table 8.1A.

Table 8.1A Estimated Carcass Weights - Grass and Rape Fed Wethers - Second Series of Trials

Grass		Rape	
Carcass No.	Kg.	Carcass No.	Kg.
555	18.6	W153	18.4
1310	18.3	W138	18.3
787	18.6	0382	17.7
569	17.7	1468	18.3
90	17.2		
975	18.8		

3. Data was analysed with computer support. Files T3R to T3W were created for the six occasions. The first four experiments compared gigots and the remaining two loins. Values of χ^2 (df4) are indicated in Table 8.2. Where the calculated value of χ^2 exceeds 9.49, 13.28 and 18.46 the probability of achieving the same results by chance alone are <0.05, <0.01 and <0.001 respectively.

Table 8.2 Values of χ^2 in Paired Comparison (Preference) and Triangle Tests: Grass versus Rape Second Trial

File T3*	Cut	N	Paired Comparisons	Triangle Tests
R	G	8	6.33	4.76
S	G	14	4.00	10.05
T	G	6	3.33	14.25
U	G	11	13.36	27.51
V	L	10	4.40	0.88
W	L	10	16.27	6.45

Values which are of statistical significance are underscored. Of the six trials, it will be noted that in two paired comparisons and three triangle tests, this applies. These values are in general agreement with those identified in more detailed studies although statistically significant results are demonstrated more frequently than in Kolmogorov-Smirnov Test. This disparity is a likely reflection of the inherent limitation of the χ^2 statistic.

Total/

Total preferences for sample A, the control, the calculated value of the z statistic and the probabilities for the grouped data are presented in Table 8.3.

Table 8.3 Paired Comparison (Preference) Tests: Grass versus Rape

File T3*	N	Cut	A preferred	Total	z	Probability<*
R	8	G	22	32	1.95	0.051
S	14	G	33	56	1.20	0.230
T	6	G	16	24	1.43	0.156
U	11	G	33	44	3.17	0.001
V	10	L	22	40	0.47	0.631
W	10	L	19	40	0.16	0.873
Totals	59		145 (61.4%)	236	3.45	0.001

*Two-tailed test. Figure in probability tables was therefore doubled. Values which are of statistical significance are underscored.

In the 4th trial file T3U, results correspond closely to those in Table 8.2 except for T3W where it is clear that preferences are evenly divided.

Similar data for the triangle tests is given in Table 8.4.

Table 8.4 Triangle Tests: Grass versus Rape

File T3*	N	Cut	Correct	Total	z	Probability<
R	8	G	15	32	1.44	0.075
S	14	G	27	56	2.22	0.013
T	6	G	15	24	2.82	0.002
U	11	G	22	44	2.82	0.002
V	10	L	14	40	5.59	0
W	10	L	16	40	6.41	0
Totals	59		107	236	3.84	0

Values which are of statistical significance are underscored.

Although the values of T3V and T3W appear to be of statistical significance, these results should be viewed with caution. The proportion of correct responses is approximately that which would have been achieved by chance alone ($33\frac{1}{3}\%$). Hence loin roasts from grass and rape fed animals must have been almost identical in flavour./

flavour. The distribution of responses of the ten subjects in these two trials is indicated in Table 8.5.

Table 8.5 Triangle Tests - Distribution of Responses in the Fifth and Sixth Trials - Files T3V and T3W

	Times Correct				
T3*	0	1	2	3	4
V	1	5	3	1	0
W	2	3	3	1	1

Tasters are usually considered to show some aptitude for tests, if the odd sample is identified on at least 50% of occasions. Whilst the number of participants on these trials is low, it can be seen that only four and five of the ten testers fall into this category. This suggests that particularly the roasted loin samples were similar irrespective of feeding regime.

Before the results of this series of trials are studied more closely, results of the Kolmogorov-Smirnov Test (which compares observed and expected cumulative frequencies of binomially distributed data) are presented in Table 8.6.

Table 8.6 Grass versus Rape - Second Series - Kolmogorov-Smirnov Test: Critical Values of D

T3*	N	Paired Comparisons	Triangle Tests
R	8	0.313	0.350
S	14	0.186	0.379*
T	6	0.300	0.600*
U	11	0.409*	0.255
V	10	0.190	0.100
W	10	0.290	0.100

*For these values of D, $p<0.05$

Retrospectively/

Retrospectively, it was a matter of regret that triangle tests were no longer used to detect possible differences in aroma. Results of tests so far considered in this chapter indicated that in some samples, differences albeit slight, existed between grass and rape fed lambs.

If it is assumed that judges identify the odd sample correctly on more than 50% of occasions, they may be considered to be selected tasters, i.e. those whose preferences are more likely to have been determined on a rational basis. Their preferences are indicated in Table 8.7.

Table 8.7* Preferences of Selected Tasters for Grass or Rape Fed Samples

T3*	N	Triangle Tests Correct 2	3	4	Total	%/N	4B	3B	3A	4A	Subjects Preferring A	B
R	8	4	2	0	6	75	0	1	2	2	5	1
S	14	9	2	0	11	79	0	3	4	1	7	4
T	6	3	3	0	6	100	0	0	2	1	3	0
U	11	2	2	2	6	55	0	0	2	3	8	0
V	10	3	1	0	4	40	0	0	2	0	5	2
W	10	3	1	1	5	50	1	2	0	2	3	6
Totals	59	24	11	3	38	64	1	6	12	9	31	13

*Note: this table corresponds to Table 6.10 in the previous chapter (6) and not to Table 6.8.

In the above table, subjects identify the odd sample correctly on 2, 3 or 4 occasions. These are the 'selected' tasters. As the difference between samples increases, the number of correct identifications and hence the number of selected tasters increases.

That flavour difference between samples is not great is indicated by the small number of subjects who identified the odd sample correctly on all four occasions. For files T3U, V and W the proportion of selected tasters is low.

If/

If the preferences of the selected tasters is considered, those subjects who preferred A or B on two of four occasions, i.e. who displayed random preference were omitted. For the remainder, either A or B was preferred on three or four occasions. Except in T3W, preference for A is demonstrated. Sample sizes are however too small to apply statistical procedures to confirm this observation. If, however, the z statistic is calculated for the total subjects in the trials where B was preferred by seven subjects and A by 21, the value is 13.36 indicating that the probability of achieving these results by chance alone is zero. These findings apply to selected tasters.

The last two columns of the table indicate the same preferences (for all tasters) as in the previous section. Again the preference is for sample A. As in the previous section, if the z statistic is calculated, such results cannot be considered to have been achieved by chance alone ($z = 21.76$).

Looking at the series of trials as a whole, it appears that there were detectable differences in the roasted gigots but not in the loins. The preference of all tasters and selected tasters appears to be for the A (grass fed) samples. The preferences of selected tasters would appear to be of greater importance. The large group of participants preferring the two samples on an equal number of occasions is of interest. This again suggests, since preference is highly subjective (demonstrated in the previous chapter), that there was little difference between samples. Preferences for A when total responses were considered was 61% (Table 8.3 of this chapter).

At the conclusion of the trials, the performance of the 22 participants was studied. Attendance at the six sessions is shown in Table 8.8.

Table 8.8/

Table 8.8 Attendance at Tasting Sessions

Occasions	6	5	4	3	2	1
Subjects	0	2	1	10	6	3

This pattern of attendance poses the problem of small panel sizes, orthogonal experimental design and inability to monitor tasters' performance.

4. The final stage in these testing procedures assessed raw and cooked samples for their aroma. Assessment of aroma had been omitted from earlier sessions as a result of discussions with participants. They had previously been dissuaded from taking part in tasting sessions because they found the procedure so distasteful and objectionable. On this occasion samples from both feeding regimes were restricted in amount and had been stored at -25°C for 33 months. It was not considered either safe or feasible to assess their flavour characteristics.

In January 1982, 43 ESCA second year BSc students of Agriculture were asked to assess the aroma of the samples following procedures indicated in Table 9.1 of the Appendix. These students had obtained statistically significant results (Chi R square 13.01 $p<0.01$) when hedonic rating scales were used to assess two samples of mature beef (steer) from different feeding regimes and one sample of cow beef. They also carried out paired comparison (difference) tests demonstrating that cow beef was tougher than steer beef. The calculated value of the z statistic was 5.49 indicating that the probability of obtaining this result by chance is zero. This group could thus be considered to show some aptitude for the sensory appraisal of meats. Their lack of experience does not seem to have handicapped performance. The design of the experiments is indicated in Table 9.1 of the Appendix.

Table 8.9/

Table 8.9 Triangle Tests to Determine Differences in Aroma Between Grass and Rape Fed Samples

	Raw Samples		Cooked Samples	
Beaker	2	AAB	1	BAA
"	3	ABA	4	AAB
"	5	BAA	6	ABA

where A = grass fed sample
B = rape fed sample

When the total correct identifications are grouped, results set out in Table 8.10 were obtained.

Table 8.10 Results of Triangle Tests - Aroma of Grass and Rape Fed Samples

	Correct	Incorrect	Total	% Correct	χ^2
Raw	62	67	129	48	11.93
Cooked	54	75	129	42	3.85

Thus the calculated values of χ^2 (1df) indicate probability levels of $p<0.001$ and $p<0.05$ respectively. Statistically significant differences in the aroma of grass and rape fed samples have been established. The validity of grouping data in this way has already been considered in Chapter 3. These results should thus be regarded with caution.

If the observed and expected frequencies of the two sets of results are compared, the calculated values of χ^2 are 37.2* and 9.43 with corresponding probability levels of $p<0.001$ and <0.05. This again suggests that, apart from demonstrating statistically significant differences between samples, these differences are greater in raw than in cooked samples. It should be noted that this test has given the same result as the grouped data in the triangle tests. The latter can thus be considered with greater confidence.

As has been previously noted, the experimenter had predicted greater ability to identify the odd sample in raw than in cooked samples. In/

In view ofthe results of the tests described above, the Wilcoxon Matched-Pairs Signed-Ranks Test for large samples (>25) was used to test this hypothesis. The value of T (the smaller sum of like-signed ranks) is 133 and of N (differences in sign of matched pairs) 26. The calculated value of z is -1.0921. This indicates that the probability of obtaining such results by chance alone is <0.1379. Thus there is no statistically significant difference in testers' performance in the two sets of trials. It should be noted that of the 43 participants, only 26 had different scores in the two trials. This test takes account of the individual responses of each taster rather than grouping results. The result is therefore more meaningful (Chapter 3). The efficiency of the test is only marginally less than the parametric t test. The median value in both series is 1.

The Kolmogorov-Smirnov Test was used to compare the observed and expected frequencies of the responses in the two trials. The data is presented in Table 8.11.

Table 8.11 Observed and Expected Frequencies in Triangle Tests to Compare Aroma of Grass and Rape Fed Samples

	Raw Samples				Cooked Samples			
Times correct	0	1	2	3	0	1	2	3
Observed frequency	9	19	8	9	12	12	15	4
Expected frequency	13	19	10	2	13	18*	10	2

*Because of rounding off, the expected frequencies did not equal 43. Thus 1 was deducted from the highest expected frequency.

The calculated values of D were 0.163 for both sample sets. For probability levels $p<0.05$ and $p<0.01$, critical values are $\frac{1.36}{\sqrt{N}}$ and $\frac{1.63}{\sqrt{N}}$ i.e. 0.207 and 0.249 respectively. Thus for both raw and cooked samples distribution of responses was binomial. This suggests that little difference existed between grass and rape fed samples. This finding is in contrast to previous results/

results and is likely to have arisen both as a result of grouping data and of the limitations of the use of the χ^2 test. The results from the Kolmogorov-Smirnov Test which take into account the frequency distribution of responses is thus likely to be more valid (Chapter 3).

5. Discussion and Conclusions

If tasters are selected on their ability to detect the odd sample correctly in the triangle test on the basis of flavour, the grass fed samples are preferred to the rape fed samples. Differences between the two feeding regimes are however so slight as to make it unlikely that the ordinary consumer would prefer lamb from a particular feeding regime. Such difference as existed in this second series of trials is obviously greater than in the first series.

In the Annual Report of the Meat Research Institute of 1974 it was noted (anon) that the feeding of rape produced no flavour defects in cooked samples. Rhodes (1971) reported flavour differences between Finnish and Scottish half bred wethers when pairs of lambs were compared. One of the pair was reared indoors without weaning. The other was removed 12-36 hours after birth and reared on cold milk ad libitum for 20 days followed by a pelleted concentrate and later on a barley-protein concentrate with hay sufficient to aid rumination. There was a highly significiant difference ($p<0.01$) in the standard error of the differences between samples. In this experiment, cooking and internal temperatures differed from those of the present study. Intensity of flavour from 'much too strong' through 'ideal' to 'much too weak' were the criteria used. Although perhaps an obvious attribute, intensity of flavour is not the only factor of importance in a study of flavour. Rhodes (1976) reported on the acceptability of lucerne treated with formaldehyde or glutaraldehyde using a tupe of hedonic scale relating

to/

to flavour, texture, juiciness and overall acceptability. Again no difference was noted. Dransfield et al (1979) indicated that no adverse comments on flavour were made on lamb which had been fed on rape. Although relating to a different species, Nute (personal communication) has found no undesirable flavour in 60 beef cattle where rape was included in the diet. Workers at MRI have thus demonstrated that whilst other feeding regimes may cause flavour defects in lamb, grazing rape does not. MRI panels can detect differences as is evident from studies quoted above.

Hence although United Kingdom results do not appear to confirm those of Park et al (1972 this does not mean that in the present trials results are invalid. Some of the reasons for different results were suggested in Chapter 6. They will be discussed more fully in Chapter 12. Others have considered either a particular flavour attribute or tasters' attitude to flavour rather than flavour per se. It was always recognised that further tests would be required to determine the nature and magnitude of any detectable differences.

CHAPTER 9

Summary - The Effect of Other Forage Crops on Lamb Flavour

In the trials described, with the possible exception of barley and cabbage fed lambs, feeding regime seems to have had little effect on the aroma or flavour of roasted lamb joints. In most trials, no link between preference and detectable flavour differences is demonstrated. Lambs from the following feeding regimes were compared:

Barley	v	Swede
Barley	v	Cabbage
Cabbage	v	Swede
Stubble (Dutch)	v	Turnips
Grass	v	Grass Silage
Grass Silage	v	Lucerne Silage

Details of carcass weights, silage composition and feeding regimes are included in the Appendix (Tables 9.2, 9.3 and 9.4)

1. The Effect of Other Forage Crops on Lamb Flavour

At the end of the trials described in Chapters 7 and 8, standard samples of joints from sheep on the other feeding regimes of the 1977 ESCA trial were available. These were from animals fed on stubble turnips, turnip swede and cabbage. In addition, following the work of Nicol and Jagusch (1971), Park et al (1972) and Park et al (1975) reporting flavour defects in lamb samples from sheep fed on lucerne, (Medicago sativa) it was intended to compare the effects of grass silage and lucerne silage on lamb flavour. This was necessary because carcasses from lucerne fed samples were not available at the time of the tests. Thus an extensive programme of investigations was planned from October 1978 onwards. For the earlier series of trials, joints from barley fed/

fed lambs were to be used as controls wherever possible. Feeding regimes and other details of lambs used in these trials are summarised below. Further details are included in Appendix Tables 9.3 and 9.4.

Breed (all trials): Scottish Blackface wethers remaining with their mothers on hill farms until August.

Weaned: In August and subsequently grazed on grass fields at lower altitudes until 10.10.77 when allocation to rape and stubble turnip feeding regimes was made. Comparisons of grass and rape fed animals were made in Chapter 6.

The feeding regimes were:

1. Stubble Turnips, i.e. 'Stubble Turnips' group

   Stubble (Dutch) turnips were grazed as the sole diet of one group of lambs from 10.10.77, i.e. from soon after weaning until slaughter (6.12.77). A small supplement of whole barley (135g/day) was introduced from 24.11.77. Remaining groups of lambs were transferred to the forage crops indicated on 5th December 1977.

2. Yellow Turnips + Whole Barley, i.e. 'Turnip' group

   Lambs were grazed on yellow turnips. These lambs were used for trials described in Chapter 7 and for comparisons of internal temperature of roasted joints by ESCA BSc students of Agriculture. Whole barley had been inroduced as a supplement (135g/day) from 24.11.77 and was gradually increased. At slaughter (12.2.78) barley consumption was 350g/day.

3. Cabbages + Whole Barley, i.e. 'Cabbage' group

   Two varieties of cabbage were grazed. Whole barley had been introduced as a supplement (135g/day) on 24.11.77 which was gradually increased until, at slaughter (4.3.78), barley consumption was 550g/day.

4./

4. Swedes + Whole Barley, i.e. 'Swedes' group

Swedes were grazed and whole barley which had been introduced as a supplement (135g/day) on 24.11.77. Supplementation of whole barley was gradually increased until, at slaughter (4.3.78) consumption was 550g/day.

5. Barley (ad libitum), i.e. 'Barley' group

This group was transferred to indoor feeding on 5.12.77. From receiving 135g/day whole barley as a supplement from 24.11.77 onwards, this group of lambs received whole barley as their sole diet from 5.12.77 onwards until, at slaughter (12.2.78) they were eating 1,130g/day.

6. Grass and Lucerne Silage Fed Lambs

Following the same weaning regime, these lambs were grazed grass aftermaths from 9th October 1977. From 9th to 21st October the silages were fed ad libitum at 9.30 and 16.30 hours. From October 21st onwards, whole barley was introduced gradually at 9.00 hours until the lambs were eating 200g/day plus silage ad libitum. Lambs were slaughtered on 8th January 1978. Further details of these silages is found in Table 9.2 of the Appendix.

7. Grass Fed Lambs

Feeding regime and rearing history are described in Chapter 8. However, to summarise these lambs remained on grass from October 5th acting as controls for rape fed lambs until slaughter on 20th November 1977.

Details of carcass weights of all lambs used may be found in Table 9.3 of the Appendix

The/

The experimental design used was the same as in the Grass versus Rape trials (second series). Table 9.1 indicates that there were six feeding regimes to be compared. For each feeding regime, either three or four trials were required. Thus it was necessary for the completion of the experimental programme that more tasters were available than in previous trials. It was thus decided to use QMC and ESCA second year BSc of Agriculture students as tasters.

In experiments which preceded the present lamb trials, extensive use was made of QMC students of Dietetics, Home Economics and Institutional Management in aroma and flavour studies. Whilst teaching programmes could not be severely disrupted, it was usually possible for student groups to participate on at least two occasions. Members of the academic and technical staff also took part in experimental programmes. Comparisons between trials could thus be made. Students who had participated previously were conscientious, scientific in their approach and eager to perform to the best of their ability. Therefore to complete investigations described in this chapter, experimental procedures were incorporated into the teaching programmes of these three student groups. In addition, ESCA students following the second year BSc course in Agriculture also assisted in these experiments. These students - often over 50 in number - showed both interest in and aptitude for the testing procedures. They had previously had an introduction to the assessment of meats by sensory appraisal techniques. As was indicated in Chapter 8 their performance was good.

In using the experimental design for the grass versus rape trials, four paired comparisons (preference), and four triangle tests in relation to flavour were required. In addition, triangle tests were used to detect possible differences in the aroma of raw and cooked samples. For the/

the assessment of flavour, individual booths with orange tungsten filament lighting were used. The aroma assessments were made in specially designated areas of the Food Science laboratory at QMC. The feeding regimes which were compared are indicated in Table 9.1.

Table 9.1 Feeding Regimes and Lamb Flavour - Comparisons from October 1978 onwards

Control Samples (A)	Code	Test Samples (B)
Barley	BS	Swede
Barley	BC	Cabbage
Cabbage	CB	Swede
Stubble Turnips (LHS)	ST	Stubble Turnips (RHS)
Grass	GGS	Grass Silage
Grass Silage	GSLS	Lucerne Silage

2. Details of each feeding regime are indicated earlier in this chapter. Pre- and post-slaughter techniques were standardised. Cooking procedure and presentation of samples were as indicated in Chapter 4. Results of these trials are indicated in the tables which follow.

Table 9.2 Values of χ^2 in Paired Comparison (Preference) and Triangle Tests: Barley versus Swede (BS)

File T3*	Cut	N	Paired Comparisons	Triangle Tests	Tasters
A	G	22	$\underline{7.77}^{a}$	$\underline{12.72}^{c}$	HE3/IM
F	L	10	$\underline{13.07}^{c}$	4.06	BA2
L	G	6	1.67	4.44	BA2
N	G	7	4.24	0.71	BA2

where HE3 = Year 3, Home Economics (Food and Nutrition Group)
IM = Institutional Management - Abridged Course
BA2 = Year 2, Home Economics
a = $p<0.10$ c = $p<0.02$

The first two groups had followed a course in Sensory Evaluation. The second year group carried out the tests during a six week module studying Meat/

Meat, Poultry and Fish. In the trial recorded in File T3A, B was preferred whilst in T3F, A was preferred. No consistent preference was thus established at this stage.

Thus where the number of tasters is high, statistically significant differences indicated by underscoring are demonstrated. Students whose results are recorded in Files T3L and T3N were from the same group. Their attendance at practical sessions was poor possibly because two members of the group were vegetarians. It was by coincidence that they were allocated to taste lamb samples from the same feeding regime.

Table 9.3 Values of χ^2 in Paired Comparison (Preference) and Triangle Tests: Barley versus Cabbage (BC)

File T3*	Cut	N	Paired Comparisons	Triangle Tests	Tasters
H	G	22	9.64[b]	8.30[a]	HE3/BA2
J	G	9	1.22	15.28[d]	BA2
K	G	17	3.86	9.80[b]	HND/DT
M	G	5	3.80	2.26	BA2

Abbreviations as in Table 9.2 plus:
HND = Year 3, Institutional Management
DT = Year 3, Dietetics
where a = $p<0.10$ b = $p<0.05$ d = $p<0.01$

It will be noted from the table above that statistically significant flavour differences in either two or three of the trials, indicated by underscoring, have been demonstrated by the use of triangle tests. In only one paired comparison was preference of statistical significance. This preference was for a joint from a cabbage fed lamb.

Table 9.4 Values of χ^2 in Paired Comparisons (Preference) and Triangle Tests: Cabbage versus Swede (CS)

File T3*	Cut	N	Paired Comparisons	Triangle Tests	Tasters
D	G	18	6.44	7.87	HND/DT
I	G	18	4.37	1.03	HND/HE3
E	L	19	2.26	3.96	HND/HE3

Abbreviations are those used in Tables 9.2 and 9.3.

Thus there were no statistically significant preferences for either sample and, as would be expected, no demonstrable differences between them.

Table 9.5 Values of χ^2 in Paired Comparisons (Preference) and Triangle Tests: Stubble Turnips LHS (A) vs. RHS (B)

File T3*	Cut	N	Paired Comparisons	Triangle Tests	Tasters
B	G	23	1.17	11.04[b]	BA2/HE3
G	G	23	10.16[b]	2.46	IM/HE3
C	L	23	4.25	2.46	HND/BA2

Abbreviations are those used in Tables 9.2 and 9.3

where b = $p<0.05$

The results in Table 9.5 are of interest. In Chapter 10, little difference is demonstrated between initial weights, total and evaporative weights and in pH values of raw and cooked samples when LHS and RHS joints are compared. In the trials where data is stored in computer files T3B and T3G respectively, a preference and a difference of statistical significance are demonstrated. When experimental records were studied it was noted that the same carcasses were involved in both trials. This is illustrated in Table 9.6 below.

Table 9.6 Source of Pooled Samples in Trials T3B and T3G

File	Cut	A Samples Animal Nos.	B Samples Animal Nos.	PC(P)	Triangles
T3B	G	436, 1361 RHS	384, 1337 RHS	NS	$p<0.05$
T3G	G	436, 1361 LHS	384, 1337 LHS	$p<0.05$	NS

The first trial had been carried out on 10.10.78 and the second on 23.10.78. Different tasters were involved on each occasion. In T3B, the odd sample was identified correctly on 42 of 92 presentations (47%). In T3G, Sample B was preferred on 52 of 92 presentations (57%). Carcass weights of each animal are indicated in Table 9.7.

Table 9.7/

Table 9.7 Carcass Weights of Stubble Turnip Fed Lambs

Carcass No.	Weight (kg)
436	15.5
1361	18.0
384	17.0
1337	19.5

In allocating samples as A or B, it can be seen that an attempt was made to cancel out the effects of differences in carcass weights. However, in doing so, differences in the pooled samples arose which were sufficient to assist identification in triangle tests and to elicit flavour preferences.

When LHS and RHS loins of the four carcasses were pooled in the trial represented by T3C, differences were cancelled out. In the paired comparisons, A was preferred on 47 of 91 occasions (52%) and in the triangle tests the odd sample was identified correctly on 34 of 92 presentations (37%). These results indicate no differences exist between samples in this trial. These findings, as indicated elsewhere in this study, demonstrate the desirability of using large numbers of carcasses from a particular feeding regime in experimental studies of this type.

For the last two trials, GGS and GSLS, tasters were ESCA 2nd Year BSc students of Agriculture. They compared roasted joints from grass and grass silage fed lambs (January 1980) and grass silage and lucerne silage fed lambs (January 1981). Their results are presented later in this chapter.

For the first four feeding regimes listed in Table 9.1, the full series of test results is available. Hence results of triangle tests on the aroma of the raw and cooked samples are presented in Table 9.8.

Table 9.8/

Table 9.8 Values of χ^2 in Triangle Tests - Aroma of Raw and Cooked Samples

Comparison	Fillet*	Cut	N	Triangle Tests (Raw)	Triangle Tests (Cooked)
BS	A	G	22	<u>12.72</u>[d]	<u>7.97</u>[b]
"	F	L	10	4.06	2.60
"	L	G	6	<u>15.94</u>[d]	1.31
"	N	G	7	4.09	1.20
BC	H	G	22	1.01	1.16
"	J	G	17	1.00	4.44
"	K	G	9	<u>10.13</u>[b]	4.75
"	M	G	5	1.75	1.07
CS	D	G	18	5.50	3.37
"	I	G	18	1.06	<u>6.25</u>[a]
"	E	L	19	2.49	<u>7.05</u>[a]
ST	B	G	23	3.17	4.29
"	G	G	23	1.41	<u>13.98</u>[d]
"	C	L	23	1.95	2.73

where a = $p<0.10$ b = $p<0.05$ d = $p<0.01$

Values of statistical significance have been underscored.

From Table 9.8, no consistent pattern of differences in aroma has been demonstrated. In three of the eight comparisons where joints from barley fed lambs were used as controls, statistically significant differences are present. In one of these three, detectable difference persists after cooking. A tendency for difference in the aroma of cooked samples may be noted when joints from cabbage and swede fed lambs are compared. The finding in relation to the two pairs of left side gigot joints from lambs fed stubble turnips is difficult to explain. There were 23 participants in this experiment and thus 69 presentations. In these presentations, the odd sample was correctly identified in 34 (49%). These figures relate to cooked samples. The corresponding figures for the aroma of raw samples were correct identifications in 20 of 69 presentations (29%) and for flavour 34 of 92 presentations (37%). The interfering factor producing a correct identification in only 29 % of presentations of raw samples may perhaps be linked to the discussion on paired comparisons and triangle tests which follows Table 9.6. In any case, there would appear to be some unusual and variable flavour and odour attributes in joints from lambs on this feeding regime particularly as variation in carcass weight was the only variable in this particular experiment. Similar findings were not obtained from the two pairs of right hand gigots. Both student groups participating in this aberrant trial had followed a College course on Sensory Appraisal.

As in the Grass versus Rape trials, the critical values of D which allow the null hypothesis to be accepted or rejected using the Kilmogorov-Smirnov Test were calculated. The rationale of this test is considered in Chapters 3 and 6. The critical values are presented in Table 9.9.

Table 9.9/

Table 9.9 Forage Crop Comparisons Kolmogorov-Smirnov Test - Critical Values of D

Comparison	File T3*	Cut	N	PCP	Triangle(F)	Triangle(AR)	Triangle(AC)
BS	A	G	22	0.236	0.105	0.332[b]	0.100
"	F	L	10	0.390[a]	0.340	0.260	0.100
"	L	G	6	0.200	0.300	0.417	0.200
"	N	G	7	0.257	0.086	0.171	0.157
BC	H	G	22	0.236	0.273[a]	0.036	0.036
"	J	G	17	0.200	0.241	0.082	0.176
"	M	G	5	0.300	0.216	0.140	0.140
"	K	G	9	0.144	0.489[b]	0.300	0.300
CS	D	G	18	0.133	0.200	0.183	0.183
"	I	G	18	0.144	0.094	0.094	0.261
"	E	L	19	0.058	0.174	0.158	0.189
ST	B	G	23	0.052	0.196	0.165	0.174
"	G	G	23	0.209	0.113	0.087	0.217
"	C	L	23	0.078	0.113	0.087	0.139

Abbreviations used are as in Table 6.5

Of the 56 trials studied, in only four were the critical values of D of sufficient magnitude to allow the null hypothesis to be rejected (for two $p<0.10$, for one $p<0.05$ and for the fourth $p<0.01$).

For T3F, Sample A was preferred in 27 of 40 presentations ($p<0.05$). The distribution was not binomial in that in seven of the ten participants preferred A on either three or four occasions where the expected cumulative frequencies of these two categories was only 3.1. No statistically significant difference in flavour was demonstrated. It was unfortunate that this group used the earlier recording form where only two triangle tests of the samples were made. Seven of the ten participants identified the odd sample on one occasion and two on both presentations. The correct identification was made on 11 of 20 presentations ($p<0.05$). The triangle test (flavour) might have produced a higher value of D had four presentations been made.

For T3H, 40 correct identifications of the odd sample were made in 88 presentations ($p<0.05$). The value of χ^2 was 8.30 ($p<0.10$). Twelve of the 22 subjects identified the odd sample correctly on 3 or 4 presentations where the cumulative expected frequencies were 6.9. This data indicates why the actual frequencies were not binomially distributed. In the paired comparisons, the calculated value of χ^2 was 9.64 ($p<0.05$). Thus a flavour difference between the two samples existed. For T3K, when the same feeding regimes were compared, the critical value of D indicated a probability of $p<0.05$, eight of the nine subjects identified the odd sample correctly on two or three occasions in triangle tests. This explains the distribution of the response pattern. Thus 21 correct responses were obtained from 36 presentations ($p<0.01$). The calculated value of χ^2 was 15.28 ($p<0.01$). This clearly established flavour difference did not affect flavour preference/

preference since whilst in both T3H and T3K, B, the cabbage fed sample was preferred, results were not of statistical significance.

For T3A, where the critical value of D indicated a probability f $p<0.05$, the value of χ^2 for the aroma of raw samples was 12.72 ($p<0.01$). Thirteen of the 22 subjects identified the odd sample correctly on two or three occasions where the cumulative frequencies were 5.7. Thus the actual frequencies were not binomially distributed.

In the remaining 52 trials, cumulative actual and expected frequencies do not differ significantly.

Total preferences for Sample A, the control, the calculated values of the z statistic and the probabilities are presented in Table 9.10.

Table 9.10 Paired Comparisons (Preference) Tests: Forage Crop Comparisons

Comparison	File T3*	Cut	N	A Preferred	Total	z	Probability< *
BS	A	G	22	31	88	2.665	0.008[d]
"	F	L	10	27	40	2.005	0.029[b]
"	L	G	6	7	16	0.250	0.803
"	N	G	7	13	24	0.204	0.854
BC	H	G	22	36	88	1.600	0.109
"	J	G	17	26	68	1.819	0.069[a]
"	M	G	5	13	20	1.118	0.407
"	K	G	9	15	36	0.833	0.131
CS	D	G	18	36	72	-0.118	0.904
"	I	G	18	40	72	0.825	0.407
"	E	L	19	41	76	0.574	0.569
ST	B	G	23	46	91	0	1.000
"	G	G	23	52	92	1.147	0.834
"	C	L	23	47	91	0.210	0.230

*Two tailed tests. Values in probability tables was therefore doubled to obtain these results.

where a = $p<0.10$ b = $p<0.05$ c = $p<0.02$ d = $p<0.01$

In/

In the Barley versus Swede (BS) trials, two results are of statistical significance. In T3A when gigots were compared, there was a statistically significant preference for Sample B (Swede) whilst in T3F, loin samples, Sample A was preferred. This may be a reflection of the highly subjective nature of preference or of difference between gigots and loins. The panel sizes when the other gigots on these feeding regimes were compared were too small for valid conclusions to be made. In any case preferences were almost equally divided. Perhaps not surprisingly when results for the four trials were pooled there were 90 preferences for Sample B out of a total of 168. This preference was not of statistical significance ($p<0.395$).

In the Barley versus Cabbage (BC) trials, none demonstrated statistically significant preference for either sample, although T3J indicates a strong preference for Sample B (Cabbage). Only in T3M was sample A preferred but since the panel size was so small these results should be viewed with caution. If the combined results for the 212 presentations are considered, Sample B was preferred in 122. This preference is of statistical significance ($p<0.033$).

In the Cabbage versus Swede trials, the Cabbage fed samples were marginally preferred in 145 of 274 presentations. This preference was not of statistical significance ($p<0.379$).

In the comparison of stubble turnips, no statistically significant preferences were demonstrated either in the individual trials or when results are considered as a whole ($p<0.363$).

Results of the triangle tests (flavour) are presented in Table 9.11. Calculated values of the z statistic and probabilities are indicated.

Table 9.11/

Table 9.11 Triangle Tests Flavour - Forage Crop Comparisons

Comparison	File T3*	Cut	N	Correct	Total	z	Probability<
BS	A	G	22	29	88	-0.188	0.425
"	F	L	10	11	20	1.818	0.0344[b]
"	L	G	4	4	16	-0.927	0.179
"	N	G	6	9	24	0.217	0.413
BC	H	G	22	40	88	2.30	0.011[c]
"	J	G	17	34	68	2.79	0.003[d]
"	M	G	5	8	20	0.395	0.345
"	K	G	9	21	36	3.005	0.001
CS	D	G	18	29	72	1.125	0.129
"	I	G	18	26	72	0.375	0.352
"	E	L	19	32	76	1.501	0.067[a]
ST	B	G	23	50	92	4.165	0
"	G	G	23	34	92	0.627	0.268
"	C	L	23	34	92	0.627	0.268

where a = p<0.10 b = p<0.05 c = p<0.02 d = p<0.01 e = p<0.001

In the four Barley versus Swede (BS) trials, there is only one result of statistical significance. Although the panel size was ten, only two triangle tests were used instead of four so that this result must be viewed with caution, particularly as only 53 correct identifications of the odd sample were made in the 148 presentations ($p<0.425$). In two of the trials, the panel size was small. Hence it seems unlikely that a detectable flavour difference existed between the two samples. The preference for the loin samples (File T3F) has a rational basis. This is not so in the case of the gigot samples (File T3A).

In the Barley versus Cabbage (BC) trials, three of the four indicate statistically significant flavour differences. In the fourth trial, there were only five participants which may account for the failure to detect differences. In the 212 presentations the odd sample was detected correctly in 103 ($p<0.0$). Whilst there are thus detectable differences in flavour, these differences are associated with preference for the cabbage fed samples only when the trial results were pooled ($p<0.033$).

In the Cabbage versus Swede (CS) trials, there was a just detectable difference in flavour only in the loin samples ($p<0.067$). In this experiment the Cabbage fed samples were preferred although this result was not of statistical significance.

The unusual result in the Stubble Turnips trials have been considered earlier in this chapter following Table 9.6. It is clear that there was a very striking difference between the flavour of the two pooled samples. This was highlighted by the χ^2 value recorded in Table 9.6.

Performance of individual tasters in paired comparisons and preference/

preference tests in each of the trials was studied. Results are presented in the tables which follow. Format is the same as in Chapter 6.

Table 9.12 Consistent Preference for A or B Samples - Forage Crop Comparisons

Comparison	File T3*	Cut	N	4A	3A	3B	4B	A	B	Total
BS	A	G	22	0	2	9	3	2	12	14
"	F	L	10	3	4	1	1	7	2	9
"	L	G	4	0	1	2	0	1	2	3
"	N	G	6	0	3	0	1	3	1	4
BC	H	G	22	1	6	8	4	7	12	19
"	J	G	17	0	2	6	2	2	8	10
"	M	G	5	1	1	0	0	2	0	2
"	K	G	9	0	2	3	1	2	4	6
CS	D	G	18	2	6	4	3	8	7	15
"	I	G	18	2	5	1	2	7	3	10
"	E	L	19	2	5	6	0	7	6	13
ST	B	G	23	1	7	4	2	8	6	14
"	G	G	23	3	4	8	4	7	12	19
"	C	L	23	3	6	4	3	9	7	16

Table 9.13 Consistent Preferences for A or B - Total for Each Feeding Regime - Forage Crop Comparisons

Regime	A	B	Total	No Preference
BS	13	17	30	12
BC	13	24	37	15
CS	22	16	38	17
ST	24	25	49	20

Table 9.13 indicates that joints from barley fed lambs was preferred less frequently than either meats from swede or cabbage fed lambs.

Samples from cabbage fed lambs were preferred to swede fed lambs. This finding is in line with the results in the BS and BC trials. Either sample was consistently preferred if the feeding regime was stubble turnips.

Performance/

Performance in triangle tests of those who consistently preferred either A or B was studied. Results are set out in Table 9.14.

Table 9.14 Performance of Consistent Subjects in Triangle Tests - Forage Crop Comparisons

Regime	Consistent	Correct>50%	%
BS	30	14	46.7
BC	37	16	43.2
CS	38	15	39.5
ST	49	27	55.1

The low proportion of those showing consistent preference for either Sample A or B who can identify the odd sample correctly in at least 50 per cent of occasions should be noted. Preference judgements - as indicated in Chapter 6 - must be viewed with considerable caution. These results may indicate why so few preferences of statistical significance were recorded in Table 9.10.

As in Chapter 6, the results can be viewed differently if preferences of selected tasters, i.e. those who identified the odd sample correctly in triangle tests in at least half the presentations, are studied. This information is set out in Table 9.15.

Table 9.15 Preferences of Selected Tasters for A and B Samples in Forage Crop Comparisons

Regime	Selected Tasters	A preferred	B preferred
BS	15	8	7
BC	23	8	15
CS	15	11	4
ST	27	15	12

The total number of selected tasters in Table 9.14 does not correspond to the total number of subjects showing consistent preference in Table 9.13 because some of the selected tasters preferred A and B on an equal number of occasions. The figures quoted in Table 9.15 tend to confirm the information in Table 9.13. Barley fed lamb remains the least preferred/

Table 9.16 Triangle Tests: Aroma of Raw Samples - Forage Crop Comparisons

Comparison	File T3*	Cut	N	χ^2	Correct	Total	z	Probability‹
BS	A	G	22	12.72[c]	34	66	3.003	0.001[e]
"	F	L	10	4.06	5	30	-2.130	0.017[c]
"	L	G	6*	15.94[d]	11	18	2.250	0.012[d]
"	N	G	7*	4.09	8	21	0.231	0.413
BC	H	G	22	1.01	21	66	-0.392	0.345
"	J	G	17	1.00	14	51	-1.040	0.419
"	M	G	5	1.75	5	15	-0.274	0.390
"	K	G	9	10.13[b]	15	27	2.245	0.013[d]
CS	D	G	18	5.50	22	54	1.010	0.156
"	I	G	18	1.06	10	54	-0.722	0.233
"	E	L	19	2.49	14	57	-1.545	0.062[a]
ST	B	G	23	3.17	29	69	1.406	0.079[a]
"	G	G	23	1.41	20	69	-0.894	0.187
"	C	L	23	1.95	27	69	0.894	0.187

where a = p‹0.10 b = p‹0.05 c = p‹0.02 d = p‹0.01 e = p‹0.001

*Vegetarian participants undertook these comparisons

preferred in relation to swede and cabbage fed lamb and cabbage fed lambs are preferred to swede fed lambs.

Aroma is an important component of flavour. Differences in the aroma of raw and cooked samples were studied using triangle tests as indicated in Table 6.1. Results of these tests are presented in Tables 9.16 and 9.17.

From Table 9.16 which relates to studies of raw samples, χ^2 values (df4) of statistical significance were demonstrated in three of the 14 trials. Two related to the aroma of the Barley versus Swede (BS) trials. When the latter are studied in detail, in three of four trials differences of statistical significance were demonstrated. It thus seems likely that in the raw samples differences in aroma existed.

In contrast to results from Tables 9.13 and 9.15 where Barley versus Cabbage (BC) results seemed to show greater differences (admittedly not always of statistical significance), in only one of four trials was a statistically significant difference in aroma established.

For Cabbage versus Swede and Stubble Turnips trials although two of the trials could be considered in some laboratories to have demonstrated statistically significant differences in aroma, in the trials described in this study, it seems unlikely that differences in the aroma of raw samples existed.

If the results in Table 9.17 are considered in four of the 14 trials (two if a value of $p<0.10$ is considered to be unacceptable) values of χ^2 indicate results of statistical significance. When total results are compared, in only one of the Barley versus Swede trials (File T3F) was a difference of statistical significance established ($p<0.026$).

Table 9.17/

Table 9.17 Triangle Tests: Aroma of Cooked Samples - Forage Crop Comparisons

Comparison	File T3*	Cut	N	X^2	Correct	Total	z	Probability<
BS	A	G	22	7.97[b]	22	66	-0.131	0.448
"	F	L	10	2.60	11	30	1.936	0.026[b]
"	L	G	6*	1.31	4**	18	-1.25	0.106
"	N	G	7*	1.20	8	21	0.232	0.413
BC	H	G	22	1.16	21	66	-0.392	0.348
"	J	G	17	4.44	14	51	-1.040	0.149
"	M	G	5	1.07	6	15	2.739	0.003[d]
"	K	G	9	4.75	14	27	1.837	0.033[b]
CS	D	G	18	3.37	24	54	1.588	0.056[a]
"	I	G	18	6.25[a]	10**	54	-2.454	0.007[d]
"	E	L	19	7.05[a]	20	57	0.141	0.433
ST	B	G	23	4.39	15**	69	-2.171	0.015[c]
"	G	G	23	13.98[c]	34	69	2.681	0.004[d]
"	C	L	23	2.73	17**	69	-1.660	0.049[b]

* Vegetarian participants undertook these comparisons

**Assumptions of triangle test procedures violated in that correct identifications occurred in less than 30% of presentations, i.e. in 22.2, 18.5, 21.7 and 24.6% respectively. These results indicate that interfering factors of an unspecified nature existed.

where a = p<0.10 b = p<0.05 c =<0.02 d =<p 0.01

For the Barley versus Cabbage (BC) trials only one result was of statistical significance ($p<0.05$). This was the same trial as in Table 9.16 where statistically significant difference in aroma of raw samples was established ($p<0.013$). In three Cabbage versus Swede (CS) trials, significant differences in aroma of cooked samples were detected in two of the trials (T3D and T3I). For these two, no differences had been detected in corresponding raw samples, whereas for T3L, where loin samples were compared, the converse applied, i.e. a marginal difference in aroma in the raw samples was not present when the joints were cooked. A similar pattern of inconsistency between raw and cooked samples arises in the Stubble Turnips trials when a marginal difference in only one raw sample ($p<0.10$) is converted to statistically significant difference in aroma of the cooked samples in all three trials.

Thus in the same trial, findings related to raw and cooked samples from the same carcass are not always in agreement. Interactions occurring between polar compounds such as structural phospholipids of intramuscular fat and compounds of lean tissue reported by Mottram and Edwards (1983) could perhaps contribute to such disparities (See Chapter 1).

Considering the results of the forage crops comparisons in greater detail, in the 135 presentations of raw samples for aroma assessment in the Barley versus Swede trials, 58 correct identifications ofthe odd sample were made ($p<0.011$). Thus statistically significant differences existed in the aroma of joints from the two feeding regimes. This finding confirms the discussion which followed Table 9.16. In the 135 presentations of the corresponding cooked samples, only 45 correct identifications of the odd sample were made ($33^{1}/3\%$). This result was not of statistical significance. Hence, as in the pork trials, substances responsible for these differences in the aroma of raw samples are likely to decompose or volatilise during the cooking process.

In/

In the Barley versus Cabbage trials in testing both raw and cooked samples, of the 159 presentations, only 55 were correct identifications of the odd sample ($p<0.401$). Thus no differences in the aroma of either raw or cooked samples for these feeding regimes was established. Grouping the data in this way indicates that the single result of statistical significance in Table 9.16 (T3K) and the two results in Table 9.17 (T3M and T3K) are likely to be misleading since panel sizes were only five and nine respectively.

When samples from cabbage fed and swede fed lambs are compared, in the 165 presentations 52 and 54 correct identifications of the odd sample were made in raw and cooked samples respectively. These results are not of statistical significance. For raw samples, this confirms the assumption following Table 9.16, i.e. that there is no difference in their aroma. In cooked samples, where some difference appears to be present in two of three trials, the probabilities computed from the z statistic are <0.056 and <0.007 for T3D and T3I respectively. However, in the former trial the probability value does not meet the stringency of the $p<0.05$ criterion whilst in the latter it has already been indicated that the assumption of triangle test procedures has been violated and that results should be viewed with suspicion. Hence it seems likely that if differences are present in the cooked samples they are so slight that they are unlikely to be detected by the majority of consumers.

In the Stubble Turnips trials, in the 207 presentations, 76 and 66 correct identifications were made in raw and cooked samples respectively. These results are not of statistical significance. As in the Cabbage versus Swede trials considered in the previous paragraph, a single trial where very slight differences were established ($p<0.10$) suggests/

suggests that little difference existed between raw samples. However differences appear to be demonstrated in the aroma of cooked samples. Possible reasons for differences are considered in discussions following Table 9.17. The most likely cause of these results is that in both Trials T3B and T3C where there were 69 presentations, the odd sample was correctly identified in only 15 and 17 presentations respectively. Because the assumption of the triangle test is thus violated, these results are suspect. One value of χ^2 (Table 9.8) was 13.98 ($p<0.01$ df 4). This indicates statistically significant differences in aroma. Thus Trial T3G where subjects identified the odd sample correctly in 34 of 69 presentations ($p<0.01$) and where the calculated value of χ^2 is 13.98 is the only trial where differences between samples has been demonstrated. In view of the anomalous results of the other two trials, it cannot thus be assumed that differences will be demonstrated in future comparisons.

Discussion of the results of triangle tests comparing the aroma of raw and cooked samples from these forage crops has made rejection of the hypothesis that differences are more readily detected in raw than in cooked samples almost inevitable. However the results of the Wilcoxon Matched-Pairs Signed-Rank Tests are indicated in Table 9.18.

Table 9.18/

Table 9.18 Wilcoxon Matched-Pairs Signed-Ranks Tests - Forage Crop Comparisons

Comparison	File T3*	N	DS	T	+	-	->+*	Probability<
BS	A	22	17	44	109	44	No	NS
"	F	10	5	2	2	13	Yes	NS
"	L	6	5	2.5	12.5	2.5	No	NS
"	N	7	6	6	15	6	No	NS
BC	H	22	16	68	68	68	=	NS
"	J	17	13	45.5	45.5	45.5	=	NS
"	M	5	4	4	4	6	Yes	NS
"	K	9	4	4	6	4	No	NS
CS	D	18	9	20	20	25	Yes	NS
"	I	18	7	3.5	24.5	3.5	No	NS
"	E	19	14	30	30	75	Yes	NS
ST	B	23	20	47.5	162.5	47.5	No	0.025
"	G	23	15	22	22	98	Yes	0.025
"	C	23	17	59.5	59.5	93.5	Yes	NS

*Minus like-signed ranks > plus like-signed ranks

Only in the Stubble Turnips trials are values of T of statistical significance. Anomalies in this series of trials have already been considered. It should however be noted that in one trial T3B that the sum of positive ranks exceeds the sum of negative ranks whereas in the other the converse applies. This result is contradictory in a one tailed test and these results should be viewed with caution. As was to be expected, it has not been shown that it is easier to identify the odd sample in triangle tests to detect differences in aroma in raw than in cooked samples.

3. During January 1980, ESCA 2nd Year BSc students of Agriculture had an introductory lecture followed by a practical session to indicate the use of sensory appraisal techniques in the assessment of eating quality of meats. Beef and lamb samples were tested. As indicated in Chapter 11, their performance in these tests was good. They showed the ability to distinguish in triangle tests between the aroma of beef and lamb samples ($p<0.001$) either with or without visible fat. They could also identify the odd samples (beef) correctly on the basis of flavour in a triangle test where lean lamb and beef were presented ($p<0.001$). These results have been considered in greater detail in Chapter 11 but are included here to provide some indication of subjects' performance.

This student group carried out the trials of roasted joints from grass and grass silage fed lambs (G GS trials) using Paired Comparison (Preference) tests. Cooking procedures and presentation of samples was as described in Chapter 4. Position of A and B samples was randomised. Appendix Table 9.1 indicates how presentations were made in both triangle tests and paired comparisons. In the latter, replicated samples were used. Preferences on the basis of both flavour and general acceptability were assessed. Results are presented in Table 9.19.

Table 9.19/

Table 9.19 Flavour Preferences in G GS Trials

Sample	Tasters	Flavour Preferred	%	General Acceptability Preferred	%
Grass (A)	61	41	67	35	57
Grass Silage B	61	20	33	26	43

Thus the flavour of the roast from the grass fed sample was preferred. This preference is of statistical significance ($p<0.005$). General acceptability is a much less precise characteristic taking into account as it does such factors as appearance, flavour and texture. The latter includes characteristics such as tenderness, juiciness and mouthfeel. Experiments at QMC suggest that tasters find this test far more difficult to carry out. The results in the present trial, where no statistically significant preference for either sample ($p<0.154$) were thus perhaps to be expected.

Since both samples in both tests were the same, tasters either selected sample A or B in both tests (consistent) or preference differed between tests (inconsistent). A summary of their results is given in Table 9.20.

Table 9.20 G GS Trials: Comparing Tasters Preferences

Sample Selected	Results of Both Tests Consistent	Inconsistent
A*	26	14
B	11	9

*Results from tasters in both tests were not always complete.

Using this data a 2 x 2 contingency table was constructed. The observed and expected frequencies were compared and the value of χ^2 calculated. Since this value is 1.577, there is no statistically significant relationship between preference for either sample and consistency/

consistency of preference ($p<0.212$).

In January 1981, a second group of ESCA 2nd Year BSc students of Agriculture compared joints from grass silage and lucerne silage fed sheep (GS LS trials). On this occasion, the animals were of comparable age and feeding regimes and were slaughtered on the same day. Post slaughter conditions, frozen storage, cooking procedures and presentation to tasters were standardised.

This group of students showed no difference in liking - assessed by hedonic scale - between young, mature and cow beef (Chi R square with 2df is 0.63). They were however able to detect that cow beef was tougher than young beef in a Paired Comparison (Difference) Test. The correct identification of the tougher sample was made on 41 of 52 occasions (79%). The calculated value of the z statistic (4.02) indicates that this result was not due to chance.

The group carried out triangle tests on the aroma of grass silage (A) and lucerne silage (B) samples with and without visible fat. Presentation was as indicated in Appendix Table 9.1 and followed the same procedures as for the first series of grass versus rape trials described in Chapter 6 expect that tubes in beakers 1, 4 and 6 contained lean only and in beakers 2, 3 and 5 lean + fat. In each trial the identical samples were A (grass silage) and the odd sample was B (lucerne silage).

There were 49 subjects and hence 147 presentations of lean and lean + fat samples. The odd sample was identified correctly on 49 ($33^{1}/3\%$) of occasions for both lean and lean + fat samples.

Each tester could identify the odd sample on zero, one, two or three occasions. Hence it is possible by comparing the observed and expected frequencies to calculate χ^2 values. These values are 4.79 and 3.03 respectively (3df) ($p<0.20$ and $p<0.50$). Hence it is confirmed that/

that there is no detectable difference between the aroma of the lean and lean + fat samples derived from grass silage and lucerne silage fed lambs.

Triangle tests were used to identify differences in flavour. The correct identification of the odd sample B was made on 21 of 52 occasions (40%). To obtain results of statistical significance ($p<0.05$), 24 correct identifications would have been required. Perhaps not unexpectedly, since aroma is a major component of flavour, no statistically significant differences were detected by triangle tests in the flavour of samples from these two feeding regimes.

Structured Paired Comparison (Preference) Tests were also used. Results are presented in Table 9.21. Tasters were requested to state their preference on the basis of flavour.

Table 9.21 GS LS Trials: Paired Comparisons

Sample Preferred		%
A	31	60
B	21	40
	N = 52	

The calculated value of the z statistic indicates that these results are not of statistical significance ($p<0.136$). This tendency for the joint from the grass silage sample to be preferred is perhaps of interest. Its validity is open to challenge in that no statistically significant differences were established in the aroma or flavour of the samples. There can thus be no rational basis for preference, although as is noted in Chapter 3, inexperienced tasters may perform better in preference than in discriminatory tests.

This particular group also compared 'conventional' pork and boar meat and were required to state their preferences for each of the samples on/

on the basis of 'General Acceptability'. For the 49 subjects, pork was the preferred sample for 31 (63%) ($p<0.08$). Their performance, whilst not perhaps as good as some other groups at ESCA is at least satisfactory.

Looking at these two trials by ESCA students, grass fed lamb was preferred to grass silage fed lamb whilst the latter was preferred to lucerne silage fed lamb. There was however little difference in the aroma or flavour of the samples when they were compared. Hence it seems likely that the average consumer would find lamb from any of these three feeding regimes equally acceptable.

4. Discussion and Conclusions

In addition to the Grass versus Rape trials described in Chapters 7 and 8, the effects of other forage crops on lamb flavour were studied. The same testing procedures were used as in the earlier trials. Details of feeding regimes and slaughter dates are indicated at the beginning of this chapter.

It has already been explained that because of reluctance of tasters in earlier trials to attend sessions on a regular basis, that these trials were incorporated into teaching programmes at both QMC and ESCA. The disadvantages of making assessments in this way were twofold. Tasters participated in only two sessions. Hence it was not possible to monitor their performance. An additional disadvantage was that since practical classes at QMC are organised in small groups, any absentees caused panel membership in some of the trials to be too small for valid conclusions to be reached. Despite their inexperience most QMC students who were selected performed well in such procedures although it could not have been foreseen that so many of them would be vegetarians!

Calculation/

Calculation of χ^2 values allowed preference for samples from a particular feeding regime to be assessed as well as detecting differences between them using triangle tests. Particularly where test groups were larger, statistically significant preferences for and differences in flavour as a result of variation in feeding regime were demonstrated. In the majority of trials, there was no difference between the aroma of samples either raw or cooked which could be demonstrated by the χ^2 test. The anomalous results in trials of joints from lambs fed on stubble turnips were identified at an early stage of the analysis of the data.

Experimental results were, as expected, binomially distributed. This was indicated by the use of the Kolmogorov-Smirnov Test. In only four of 56 tests was a different distribution obtained. Possible reasons were discussed in context following Table 9.9.

Flavour preferences were studied in greater detail. From grouped results of replicated trials, calculation of the z statistic enabled the actual probability of the result in each test arising by chance alone to be considered. In only the Barley versus Swede (BS) trials were two statistically significant preferences for joints from either feeding regime demonstrated. It should be noted that on one occasion the preference was for samples from the barley fed regime and on the other for the swede fed samples! This observation suggests that, in line with other trials, preference for a particular feeding regime is either random or very slight when data is grouped in this way.

If comparisons of preference are made, barley fed lamb is less preferred than swede or cabbage fed samples. When the latter are compared, cabbage fed lamb is preferred to swede fed lamb. The problems in the Stubble Turnips regime have already been highlighted. It would seem that in combining samples - as has been indicated - that instead

of/

of achieving uniformity, the result was to produce samples which differed considerably.

There was little difference between the aroma of raw and cooked samples on any of the feeding regimes and, except on the occasions indicated above, detection of the odd sample in the triangle tests was no easier in raw than in cooked samples. There is however some evidence of statistically significant flavour differences particularly in the Barley versus Cabbage trials. ESCA tasters preferred the flavour of grass fed to grass silage fed samples. Feeding regime was not the only variable in this experiment since the grass fed lambs were slaughtered at the end of November 1977 whereas the silage fed lambs were slaughtered early in January 1978. Disparities in age of animal at slaughter of this magnitude do not appear to affect results presented elsewhere in this study. This preference was not of statistical significance when the students used the criterion 'general acceptability' where attributes other than flavour are taken into account. The following session ESCA students were unable to demonstrate preference for samples from either grass silage or lucerne silage in relation to flavour. This finding was not unexpected in that triangle tests had demonstrated no aroma differences between samples from the two feeding regimes.

In the trials described in this chapter, with the possible exception of the Barley versus Cabbage comparisons, feeding regime appears to have exerted little effect on flavour including its component aroma. Detectable differences - as identified by the well recognised triangle test - have been demonstrated on few occasions. Whilst preference is a highly subjective characteristic there seems to be no link between preference and detectable difference except in the Barley versus Cabbage/

Cabbage trials.

The series of experiments has only pointed the way to future investigations when it would clearly be highly desirable to recruit a larger number of tasters and to have more samples from each regime available for comparison. Lamb appears to be a highly variable meat which relatively inexperienced subjects find extremely difficult to assess.

CHAPTER 10

Objective Data

Summary

The presentation of data is explained. Differences in the weights of gigots and loins of standard location on left and right sides are evident but they are not of statistical significance. There are statistically significant differences both in the total and evaporative weight losses between gigots and loins cooked by standard techniques. pH values have been studied. Perhaps the most important finding is that there are statistically significant differences in both raw and cooked pH values of gigots and loins. Although these differences exist between gigots and loins, changes in pH induced by the cooking process do not differ in the two cuts. Left and right sides of the same carcass of standard anatomical location show no pH differences of statistical significance.

1. Collection of Objective Data

Factors which influence weight losses in roasted meats and pH of raw meats have received extensive study. The former are comprehensively reviewed by Harries et al (1960) in relation to beef.

Weight losses were recorded routinely whilst investigating the eating quality of beef. Standard cooking procedures were used. Findings were published to provide possible explanations of the anomalies in the literature.

As a result of the present series of trials, data is available for 166 lamb joints. Gigots and loins of standard anatomical location from the left and right sides of each animal were used. Animals had been on different feeding regimes with a mimimum of four from each regime./

regime. Data is not available for all joints used in the testing procedures. Preparation of large numbers of joints on some occasions made collection of such data impracticable.

The present information includes:

Raw and cooked weights of each joint (on the bone)

Total, drip and evaporative weight losses as above

The pH values of raw and cooked joints as above

The information was collected as a "spin-off" from the main study, i.e. the effect of feeding regime on lamb flavour. The original aim was to contribute to fundamental knowledge of the characteristics of raw and cooked lamb joints.

The possible comparisons between pH values, weight losses and sensory characteristics of the joints studied are extensive. To carry out the detailed analyses required, particularly correlations between objective parameters and results of sensory appraisal techniques, would be, in itself, a major exercise. Hence in the present study it is proposed to highlight only the major findings.

A more extensive study of the data will be made when this thesis is otherwise complete. Some of the information is included in other chapters, e.g. in Chapter 4 Establishing a Methodology, when the practice of standardising anatomical location of joints is described. The experimenter is fully aware of the importance of utilising and analysing all the data collected and of publishing findings of importance. (To be reported in Developments in Meat Science: Volume III - Ed. R.A. Lawrie Applied Science Publishers)

2. Preliminary Examination of the Objective Data

Whilst simplifying the preliminary sorting of the data by preparing only /

only frequency distributions it was considered important to identify each of the joints within a particular category. Future study of the data will thus be facilitated. Examination of the data was made under the following headings:

1. A comparison of the differences between:

   (a) Raw (on bone) weights, cooking losses - total and evaporative - and pre- and post- cooking pH values of left and right gigot and loin joints of standard anatomical location.

   (b) Gigots and loin joints, irrespective of left or right origin and of feeding regime in relation to:

      Total cooking losses

      Evaporative cooking losses

      Initial and cooked pH values.

Whilst (a) and (b) can be separated according to feeding regime and anatomical location, it is not considered that valid conclusions may be reached from a study of only four carcasses from a particular feeding regime.

Examination of the data indicates that further experimentation will be required to determine the effect of feeding regime - if any - on the parameters studied.

3. Explanation of the Tabulation of the Data

1. Differences in weight of left (LHS) and right (RHS) joints of standard anatomical location are expressed as plus or minus values in g between LHS and RHS, eg.

   *(i) LHS 643g RHS 627g Difference +16g

   (ii) LHS 536g RHS 538g Difference -2g

*These/

*These are illustrative examples only. Disparities of much greater or lesser magnitude are observed.

2. A similar practice has been followed when comparing total and evaporative losses of LHS and RHS joints of standard anatomical location from the same carcass. Weight losses are expressed as a percentage of the initial weight while joints are 'on the bone'.

3. Total and evaporative weight losses and pre- and post cooking pH values are tabulated for gigot and loin joints irrespective of feeding regime; further investigation will be assisted as indicated by identifying joints in each category. Since much of the material represented by weight loss from a joint during roasting is later collected as drip comprising exuded fat and juices, the actual yield should take account of this collected drip as well as the weight of the cooked meat. Hence the only real loss is probably evaporative (Harries et al, ibid). For this reason only total and evaporative losses have been studied at this stage in the analysis of the data.

4. In relation to pH values, raw and cooked pH values for each joint are compared. Left and right gigots and loins from the same carcass are also compared. The net changes in pH induced by the cooking process have been studied in relation to LHS and RHS joints of standard anatomical location,

eg. Change in pH (pre- versus post- cooking)

	LHS	RHS	Net change (L v R)
(i)	0.25	+0.30	-0.05
(ii)	0.20	-0.10	+0.30

The examples above are intended only to demonstrate how data relating to these changes has been analysed.

4./

4. The Initial Weights of the Joints

There are records available for 86 gigots and 80 loins. Arithmetic means and standard deviations were calculated for the 4 groups of joints. Results are given in Table 10.1.

Table 10.1 Initial Weights of Lamb Joints

Cut	LeG	RG	LeL	RL
N	45	41	43	37
Mean	607	610	631	643
Standard Deviation	85.7	84.6	122.1	117.2

(Weights have been rounded off to the nearest g) where Le = left, G = gigot, R = right, L = loin.

From Table 10.1, the arithmetic means suggest that corresponding joints from left and right sides are of remarkably similar weight. Standard Deviations suggest that this may not be so. Preliminary study of LHS and RHS joints indicates that there is considerable variation between them. Even the summary in Table 10.2 does not indicate the magnitude of these differences. This is demonstrated by the Frequency Distribution Table 10.3.

Table 10.2 Comparison of Initial Weights

	Gigots	Loins
Same weight	0	0
+/- 5g	3	4
Left > Right	23	20
Left < Right	15	11
Totals	38	31

When the data is grouped in this way, there is no statistically significant difference between left and right sides as indicated by the/

the Median Test. The wide range of differences should however be noted: there are 20 and 16 categories of difference between gigots and loins respectively.

Table 10.3 Frequency Distribution Table - Left v Right Joints (Same carcass)

Differences g	Positive Values		Negative Values	
	Gigots	Loins	Gigots	Loins
0-4	3	4		
5-9	4	2	3	
10-14	2	2	3	1
15-19	3	6	2	1
20-24	2			1
25-29	3	1	1	2
30-34		1	2	1
35-39		1		1
40-44				2
45-49	1		1	
50-54	2	1	1	1
55-59	1			
60-64	1			
65-69			1	
70-74	1		1	
75-79				
80-84				
85-89				
90-94				
95-99				
Totals	23	20	15	11

Gigots N = 38

Loins N = 31

Figure 10.1 presents this information in the form of a bar chart.

It is thus clear that whilst the weight of joints considered solely as gigots or loins appear to show uniformity as reflected by/

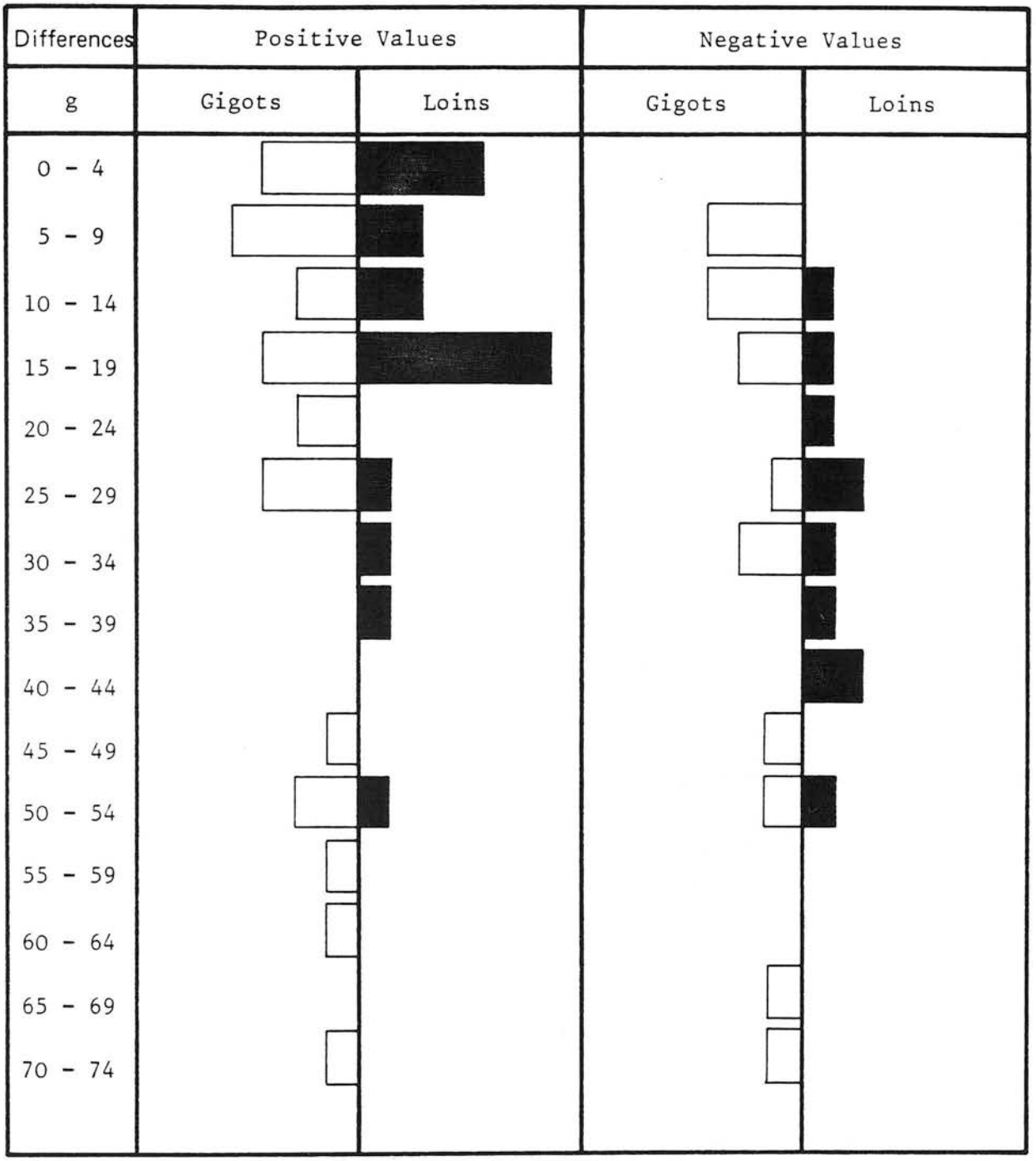

Figure 10.1 Frequency Distribution Table - Weight Differences between Left and Right Joints from same carcass

by mean weights, this does not highlight differences existing both within and between carcasses in respect of right and left sides. Predictably, these initial differences are paralleled when total and evaporative weight losses are studied.

5. Total Weight Losses of Joints

If total cooking losses from left and right sides are considered (Table 10.4), both gigots and loins appear to be very similar. Cooking losses are expressed as percentages of raw (bone in) weights.

Table 10.4 Total Cooking Losses of Gigots and Loins

Cut	LeG	RG	LeL	RL
Mean	29.3	29.1	17.0	16.7
Standard Deviation	5.0	4.5	3.7	4.3
N	45	41	43	37

(Abbreviations are as in Table 10.1)

When left and right gigots/loins from the same carcass are compared, (Table 10.5) this overall result can be seen to be misleading. When total cooking losses are compared, considerable variation within the same carcass can be demonstrated.

Table 10.5 Total Weight Losses - Left v Right Sides

Cut	Gigots	Loins
Equal	1	3
L > R	25	20
L < R	12	8
N	38	31

The distribution of these weight losses is indicated in Table 10.6 and Figure 10.2.

Tables 10.7 and 10.8 and Figure 10.3 demonstrate the total weight losses/

losses - irrespective of side or feeding regime - for gigots and loins. These comparisons indicate why the mean values and standard deviations quoted in Table 10.4 correspond so closely. There is no statistically significant difference between gigots and loins in this respect. Perhaps the most important point to emphasise is that for all experiments of this type, sample sizes should be as large as possible, since apparently similar sheep meat samples demonstrate inherent variability in relation to total cooking losses.

Table 10.6 Comparing Total Percentage Weight Losses - Left v Right Sides (Irrespective of Feeding Regime)

	% Loss (Total)	Gigots		Loins	
Minus Values	10-11			1	
	9-10			1	
	8-9				Gigots N = 38
	7-8	2			Loins N = 31
	6-7	2			
	5-6	1			
	4-5	1		1	
	3-4	3	(31% of total)	3	(26% of total)
	2-3	3		2	
	1-2				
	1-0				
Zero	0	1		3	
	0-1	4		4	
	1-2	4		3	
	2-3	5		4	
	3-4			3	
	4-5	2		2	
	5-6	2		3	
	6-7	4			
	7-8	3			
	8-9	1	(66% of total)		(68% of total)
	9-10				
Plus Values	10-11				
	11-12			1	

Figure 10.2 presents this information in the form of a bar chart.

It will be observed (Table 10.7) below, that total weight losses fall into only five categories for both gigots and loins. The percentage/

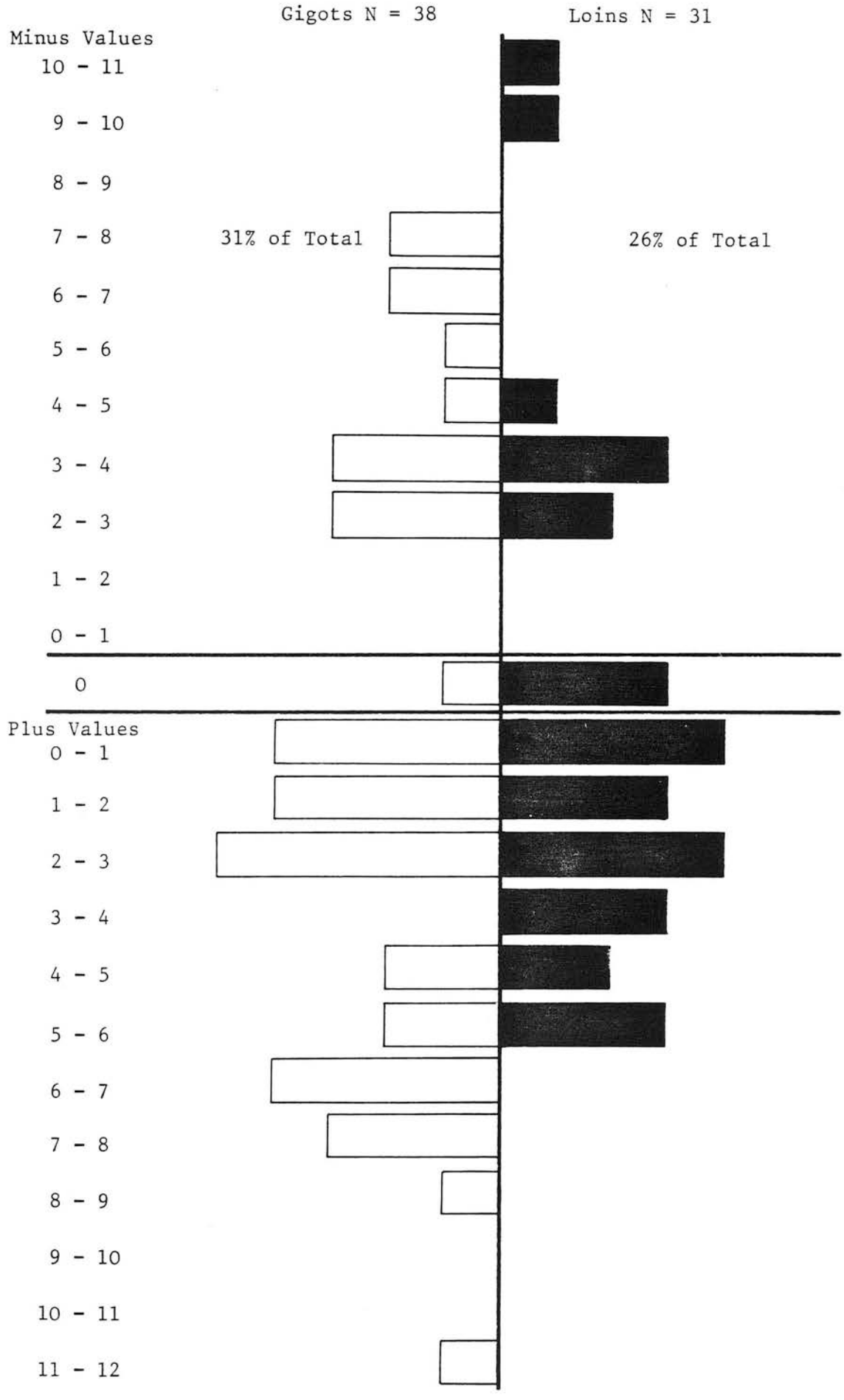

Figure 10.2 Comparing Total Percentage Weight Losses - Left Versus Right Sides (Irrespective of Feeding Regime)

percentage of values falling into the two major categories for gigots and loins are 72.1 and 76.3 respectively. Although these figures appear to be in good agreement, they represent the categories 25-34% and 10-19% total weight losses for gigots and loins respectively.

Table 10.7 Total % Weight Losses Irrespective of Feeding Regime Gigots and Loins

% Losses Total	5-9	10-14	15-19	20-24	25-29	30-34	35-39
Gigots	0	0	2	11	38	23	12
Loins	3	22	39	13	3	0	0

Frequency distributions of the total percentage weight losses in gigots and loins, irrespective of feeding regime, are indicated in Table 10.8 and Figure 10.3 below.

Table 10.8 Total Weight Losses in Gigots and Loins Irrespective of Feeding Regime

% Loss	Gigots Frequency	Cumulative Frequency	Loins Frequency	Cumulative Frequency
0-4	0	0	0	0
5-9	0	0	3	3
10-14	0	0	22	25
15-19	2	2	39	64
20-24	11	13	13	77
25-29	38	51	3	80
30-34	23	74	0	80
35-39	12	86	0	80

The great difference in the frequency distribution between gigots and loins is even more striking when the bar chart is studied (Figure 10.3).

It was thus decided to use the Median Test to determine if differences/

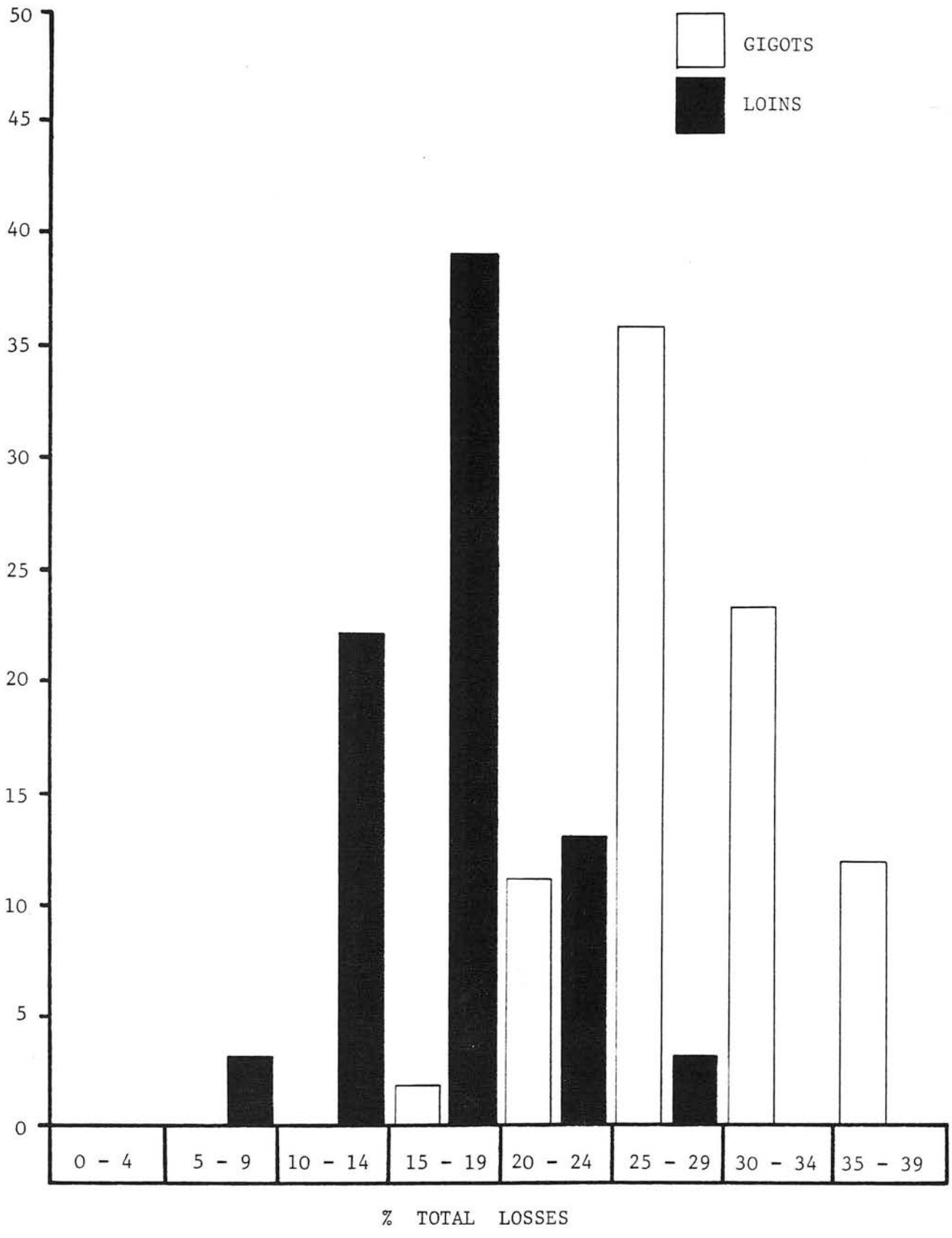

Figure 10.3 Total weight losses - Frequency Distributions: Gigots and Loins

differences in the frequency distributions of gigots and loins was of statistical significance.

Table 10.9, indicates the Cumulative Frequencies for gigots and loins separately. To calculate the Common Median it is necessary to combine this data for gigots and loins.

Table 10.9 Cumulative Frequencies - Total Weight Losses

Gigots and Loins

% Loss	Cumulative Frequency Gigots	Cumulative Frequency Loins	Sum (G + L) Cumulative Frequencies
0-4	0	0	
5-9	0	3	3
10-14	0	25	25
15-19	2	64	66
20-24	13	77	90
25-29	51	80	131
30-34	74	80	154
35-39	86	80	166

From the table above it is clear that the Common Median is to be found in the 20-24% weight loss category (N = 24). The mean of the 83rd value (22.3) and the 84th value (22.6) is 22.45. This is therefore the Common Median. Within this category, nine gigot and nine loin joints were below the Common Median with two gigotsand four loins above it.

Values were then dichotomised by assigning a plus sign to values exceeding the Common Median, i.e. all values in the 25-39% range and a minus sign to values at or below the Common Median, i.e. all values 5-19% range. Plus and minus signs were allocated within the 20-24% as/

as indicated in the previous paragraph. A 2 x 2 array was then prepared. The value of χ^2 was calculated according to the formula

$$= \sum_{i=1}^{r} \sum_{j=1}^{K} \frac{(Oij - Eij)^2}{Eij} \qquad \text{(Formula 10.1)}$$

where Oij = observed number of cases categorised in the ith row of jth column

Eij = number of cases expected under the null hypothesis to be categorised in the ith row of jth column.

(Observed frequencies are in bold letters in each cell and expected frequencies in fine letters in the table). Relating to Total Weight Losses, details are as follows:

	Gigots	Loins	Totals
Plus Values	42.5 **75**	39.5 **7**	82
Negative Values	43.5 **11**	40.5 **73**	84
Totals	86	80	166

Values of $\frac{(O-E)^2}{E}$ for each cell were summed to give a calculated value of $\chi^2 = 101.95$ with df $(K-1)(r-1) = 1$.

This value of χ^2 indicates a probability of <0.000. The null hypothesis is thus rejected. A statistically significant difference in the total weight losses of gigots and loins has been clearly demonstrated in this series of trials. The importance of standardising the cuts of lamb used in this type of trial is emphasised and has implications both for future experiments and for assessing experiments carried out in other laboratories.

Although of less validity in this experiment, a K x 2 table was also constructed. To meet the criteria for carrying out a χ^2 test,

some/

some grouping of the data was required. The calculated value of χ^2 confirmed the findings of the Median Test.

Grouping the data in this way i.e. at 5% intervals, whilst demonstrating striking differences between gigots and loins, could give a misleading impression of uniformity within a particular category. This point is illustrated by studying the values for gigots (N = 38) in the 25-29% category and for loins (N = 39) in the 15-19% category. For these 38 gigots where are 50 values from 25.0 through to 29.9, 28 categories are represented. Twenty two categories have no assigned values. Within the 28 categories only one had a frequency of three and nine had a frequency of two. For the 39 loins 30 of the 50 categories are represented. Within these 30 categories only two had a frequency of three and five a frequency of two. Thus within a particular category there is a considerable scatter of values.

6. Evaporative Weight Losses Irrespective of Feeding Regime - Gigots and Loins

Evaporative weight losses were obtained by subtracting the weight of drip from the total weight loss of each joint. The evaporative loss was then expressed as a percentage of the total raw (on bone) weight of each joint.

As with total weight losses, when left and right sides are considered (Table 10.10) left and right gigots and loins appear to be very similar.

Table 10.10 Evaporative Losses of Gigots and Loins

Cut	LeG	RG	LeL	RL
Mean	24.7	24.0	12.3	11.7
Standard Deviation	5.0	4.8	2.9	3.8
N	45	41	43	37

(Abbreviations are as in Table 10.1)

However/

However, considerable variation within the same carcass is demonstrated. Table 10.11 indicates the comparative losses between left and right gigots and loins. These variations are shown more clearly in Figure 10.4. It is again evident that, because of inherent variability experimental findings should be based on large sample sizes. There is no statistically significant difference between left and right gigots and loins.

Evaporative losses, irrespective of side or feeding regime are indicated in Table 10.12 and Figure 10.5.

Table 10.11 Comparing Evaporative Weight Losses - Left versus Right Sides (Irrespective of Feeding Regime)

		Gigots N = 38		Loins N = 31	
		% Loss (Evaporative)			
Minus Values	13-12	1		1	
	12-11				
	11-10	1		1	
	10-9				
	9-8	3		1	
	8-7	2			
	7-6	1	47% of total		42% of total
	6-5				
	5-4			1	
	4-3	3		3	
	3-2	1		1	
	2-1	2		3	
	1-0	3		2	
	0	1		2	
	0-1	1		5	
	1-2	3		5	
	2-3	4		1	
	3-4	3		2	
	4-5	2	50% of total	1	52% of total
	5-6			1	
	6-7	3			
	7-8	1			
	8-9				
Plus Values	9-10	1		1	
	11-12	1			

Figure 10.4 presents this information in the form of a bar chart.

Table/

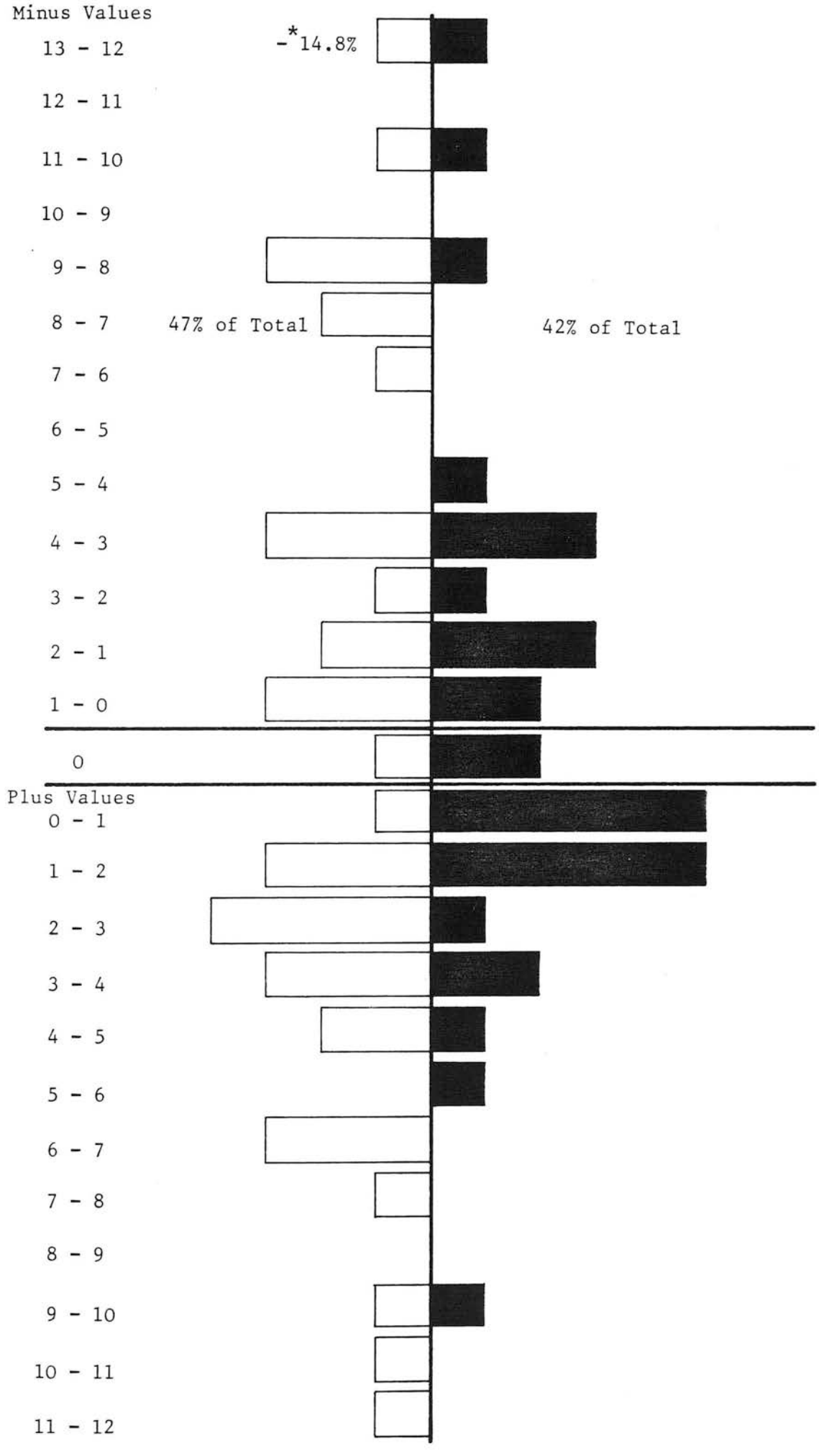

Figure 10.4 Comparing Evaporative Weight Losses - Left Versus Right Side (Irrespective of Feeding Regime)

Table 10.12 Frequency Distribution: Evaporative Weight Losses

Gigots and Loins

% Evaporative Loss	Gigots	Loins	Cumulative Total
0-4	0	3	3
5-9	0	10	13
10-14	7	58	78
15-19	8	6	92
20-24	30	3	125
25-29	31	0	156
30-34	10	0	166
N	86	80	

The Common Median was thus found in the 15-19% category where N = 14. Examination indicates that the mean of the 83rd and 84th values is 16.95. This represents the Common Median. As previous indicated, within this category the values were assigned as below:

	Plus	Minus	Total
Gigots	7	1	8
Loins	2	4	6

As with the total weight losses, a 2 X 2 array was constructed and the value of χ^2 calculated (Formula 10.1)

	Gigots	Loins	Totals
Plus Values	$^{43}\underline{78}$	$^{40}\underline{5}$	83
Minus Values	$^{43}\underline{8}$	$^{40}\underline{75}$	83
Totals	86	80	166

When the $\frac{(O - E)^2}{E}$ values for each cell were summed the value of χ^2 was 118 (1df). The null hypothesis must therefore be rejected on the basis that $P<0.000$. Thus as with total weight losses, a statistically significant difference in evaporative weight losses between gigots and loins/

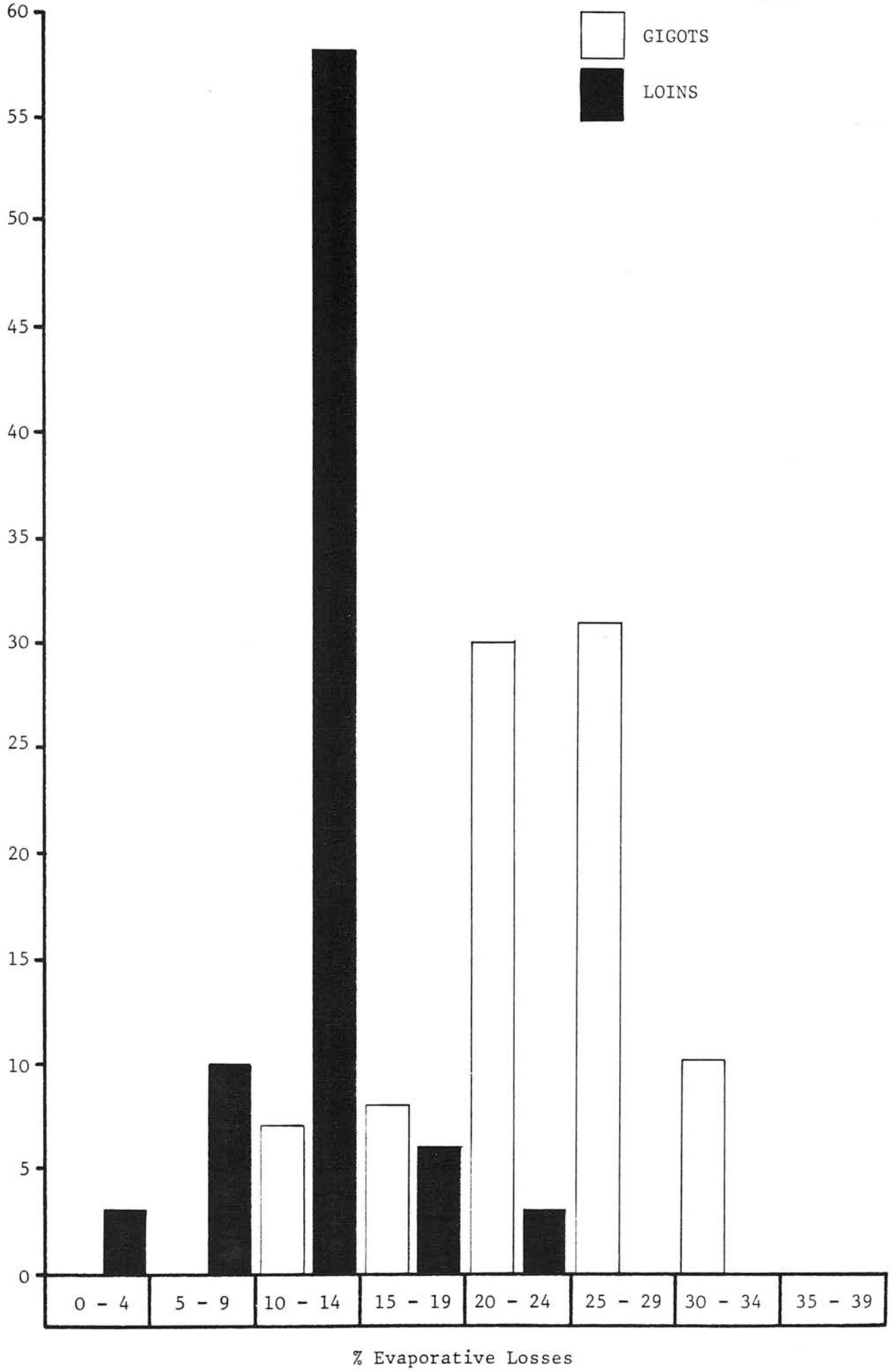

Figure 10.5 Evaporative Losses - Frequency Distributions: Gigots and Loins

loins has been demonstrated. It is recognised that grouping the data in this way, as has been indicated, conceals the variation within categories. From Table 10.12 it will be noted that evaporative losses for both gigots and loins fall into five categories. In gigots, the category for maximal evaporative loss is in 25-29% (N = 31) range: This accounts for 36% of the sample. Whereas for loins, maximal evaporative loss falls in the 10-14% range. This accounts for 72.5% of the sample. In comparing these results with those for total weight losses (Table 10.8), it can be seen that the figures for the 25-29% category for gigots are 44% and in the 15-19% for loins 49% of the total samples respectively. This suggests greater uniformity of evaporative weight losses in loins than in gigots. This difference is only 5% in respect of total weight losses. If the studies of the data are extended, weight losses both total and evaporative - suggest that loins are more homogeneous than gigots.

7. Studies of pH Values

Initial pH values

(i) As indicated earlier in this chapter, pH values of raw and cooked joints were recorded. The frequencies of the initial, i.e. raw pH values are indicated in Table 10.13 and indicated in the form of a bar chart in Figure 10.6. The considerably lower pH values for the loin roasts is very obvious. The values quoted are for all gigots and loins irrespective of feeding regime or side of the carcass.

The Median Test was carried out with plus signs being allocated to pH values above the Combined Median and minus signs to values at or below. The Combined Median i.e. the mean of the 83rd and 84th values is 5.625. The calculated value of X^2 (derived from a 2 x 2 array is 44.25 (1 df). The probability of this value being achieved by chance alone/

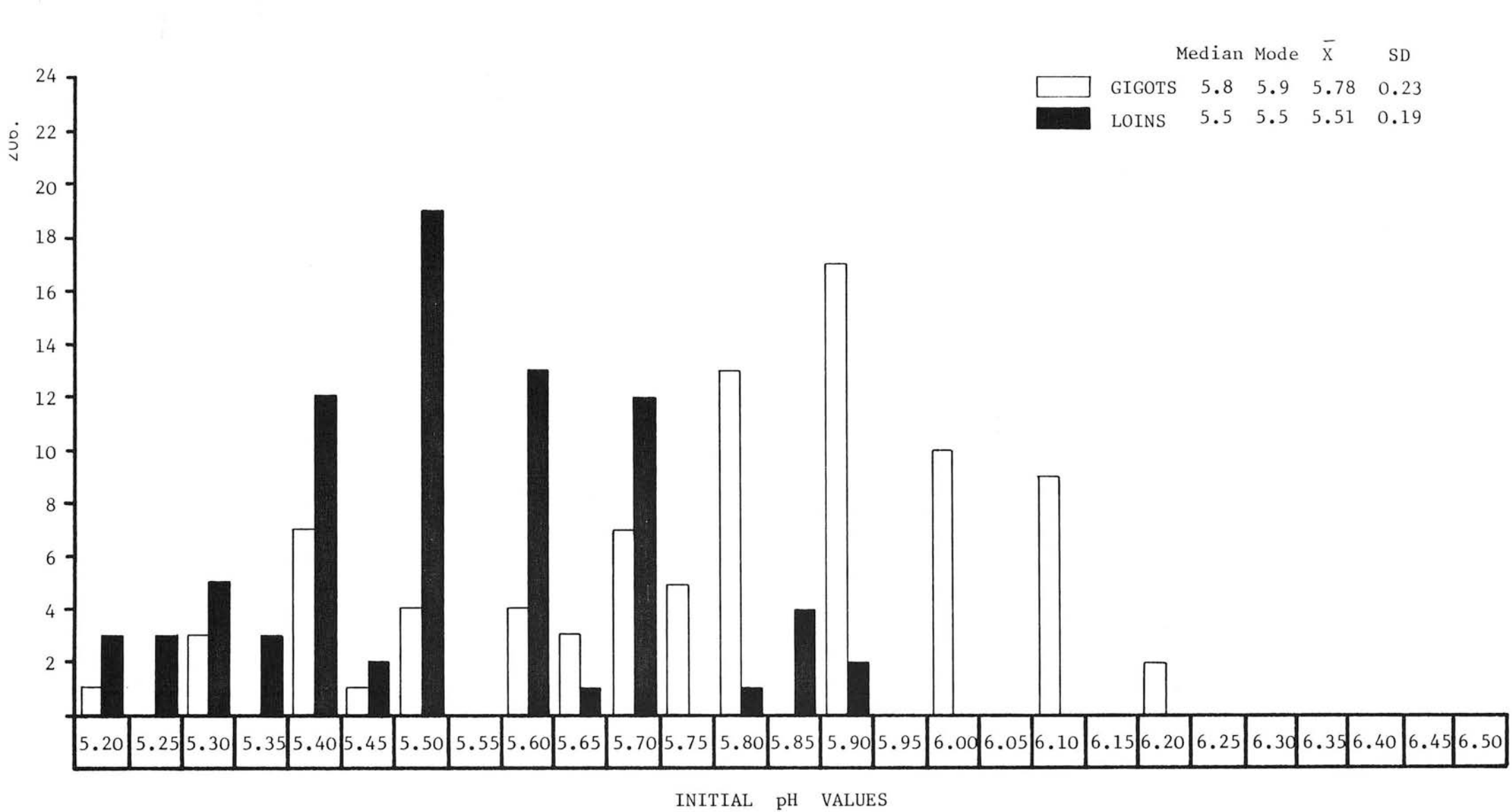

Figure 10.6 Frequency Distributions: Initial pH Gigots and Loins

alone is zero. Hence there is a statistically significant difference in the initial pH of gigots and loins. There is also less variation in loins than gigots.

Table 10.13 Initial pH Values of Gigots and Loins: Frequency Distribution

pH Value	Gigots	Loins
5.20	1	3
5.25	0	3
5.30	3	5
5.35	0	3
5.40	7	12
5.45	1	2
5.50	4	19
5.55	0	0
5.60	4	13
5.65	3	1
5.70	7	12
5.75	5	0
5.80	13	1
5.85	0	4
5.90	17	2
5.95	0	
6.00	10	
6.05	0	
6.10	9	
6.15	0	
6.20	2	
Mean	5.78	5.51
SD	0.23	0.19

That there is a statistically significant difference between gigots and loins is confirmed when the initial pH values of the 38 left and right gigot pairs and 31 loin pairs are studied.

	Gigots	Loins
Median	5.8	5.5
Mode	5.9	5.1
N	76	62

Whilst there is no significant pH difference in gigots or loins from left and right sides, when all gigots and loins are compared, the/

the Combined Median (data available but not tabulated here) is 5.625 and the calculated value of χ^2 39.76 (1 df). With this value of χ^2, results such as these cannot have been achieved by chance alone.

(ii) Cooked pH values

The cooked pH values of gigots and loins are indicated in Table 10.14 and Figure 10.7.

Table 10.14 Cooked pH Values Gigots and Loins: Frequency Distribution

pH Value	Gigots	Loins
5.50	0	0
5.55	0	1
5.60	2	17
5.65	0	2
5.70	9	19
5.75	0	5
5.80	9	11
5.85	4	2
5.90	5	13
5.95	1	0
6.00	18	6
6.05	0	0
6.10	9	1
6.15	5	0
6.20	9	1
6.25	1	1
6.30	7	1
6.35	1	
6.40	5	
6.45	0	
6.50	1	
Mean	6.03	5.75
SD	0.22	0.16

As might have been expected, as a result of protein denaturation and other reactions taking during the cooking process, in the majority of roasts/

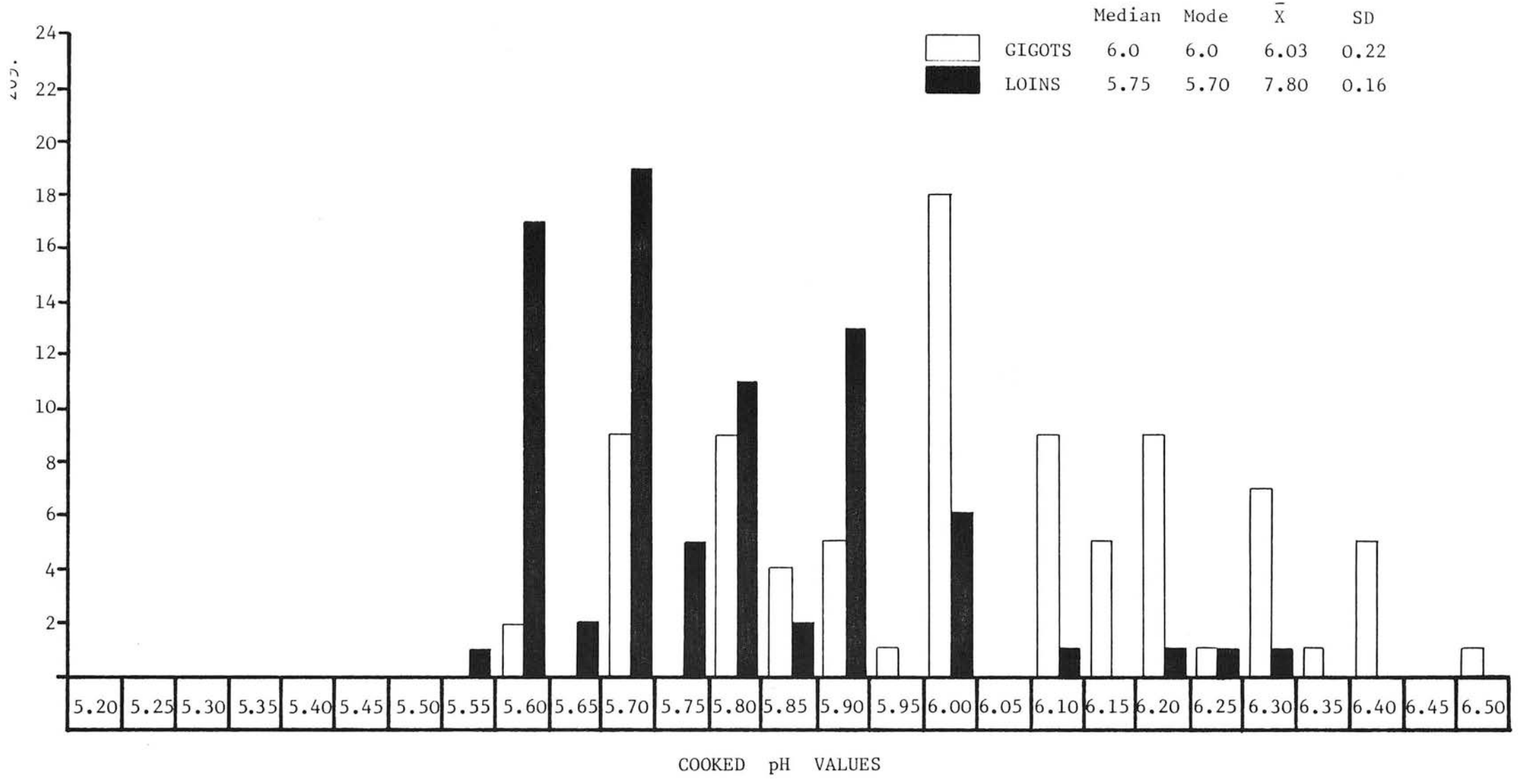

Figure 10.7 Frequency Distributions: Cooked pH Values - Gigots and Loins

roasts the cooked pH values were higher than the initial pH. Where this was not so, the pH meter was recalibrated and the pH of the slurry checked. On no occasion did the second reading alter from the first. It can only be concluded that, in some muscle tissue, changes occurring during the cooking process cause the pH to fall. (Initial pH values were always measured twice).

In the Median test, the Combined Median was 5.9 and the calculated value of χ^2 was 48.53 (1df). As with initial pH, there is thus a statistically significant difference between the pH values of gigots and loins. Loins again show lower pH values and less variation than gigots. This result is confirmed when figures for the 76 left and right gigots and 62 loins are compared. Data is available (but not tabulated here) which indicates the following for the cooked pH values.

	Gigots	Loins
Median	6.0	5.725
Mode	6.0	5.6/5.7
N	76	62

The Combined Median is 5.825 and the calculated value of χ^2 33.88 (1 df). Again, for this subgroup, a statistical significant difference has been established between pH values of cooked gigots and loins, although there is no difference between left and right sides.

(iii) Alterations in pH Values During the Cooking Process

These alterations in pH are indicated in Table 10.15 and Figure 10.8.

Table 10.15 Alterations in pH Values During the Cooking Process - Gigots and Loins : Frequency Distribution

pH Values	Minus	Gigots	Loins
0.20	"	2	1
0.15	"	0	0
0.10/			

Table 10-15 (Contd)

pH Values	Minus	Gigots	Loins
0.10	"	1	1
0.05	"	0	0
0.00	No change	2	1
0.05	Plus	2	3
0.10	"	12	11
0.15	"	2	3
0.20	"	16	16
0.25	"	7	2
0.30	"	21	17
0.35	"	5	8
0.40	"	13	14
0.45	"	1	0
0.50	"	1	3
0.55	"	0	
0.60	"	1	
N		86	80

In the Median test the Combined Median was 0.30. The calculated value of χ^2 was 1.10 (1df) $p<0.30$. Hence there is no statistically significant differences in the pH changes occurring in gigots and loins. This is despite statistically significant lower pH values in raw and cooked loins.

The changes in the loin roasts are on this occasion - as to be expected - only slightly more uniform. It should be noted that only 3 gigots and 2 loins had lower post-cooking pH values, with 2 gigots and 1 loin roast with a pH value which remained constant.

(iv) Alterations in pH During the Cooking Process - Left and Right Gigots and Loins

The alterations in left and right gigots and loins are considered separately. As in the previous section initial pH values were not taken into account. The values are quoted in Table 10.16 and Figure 10.9/

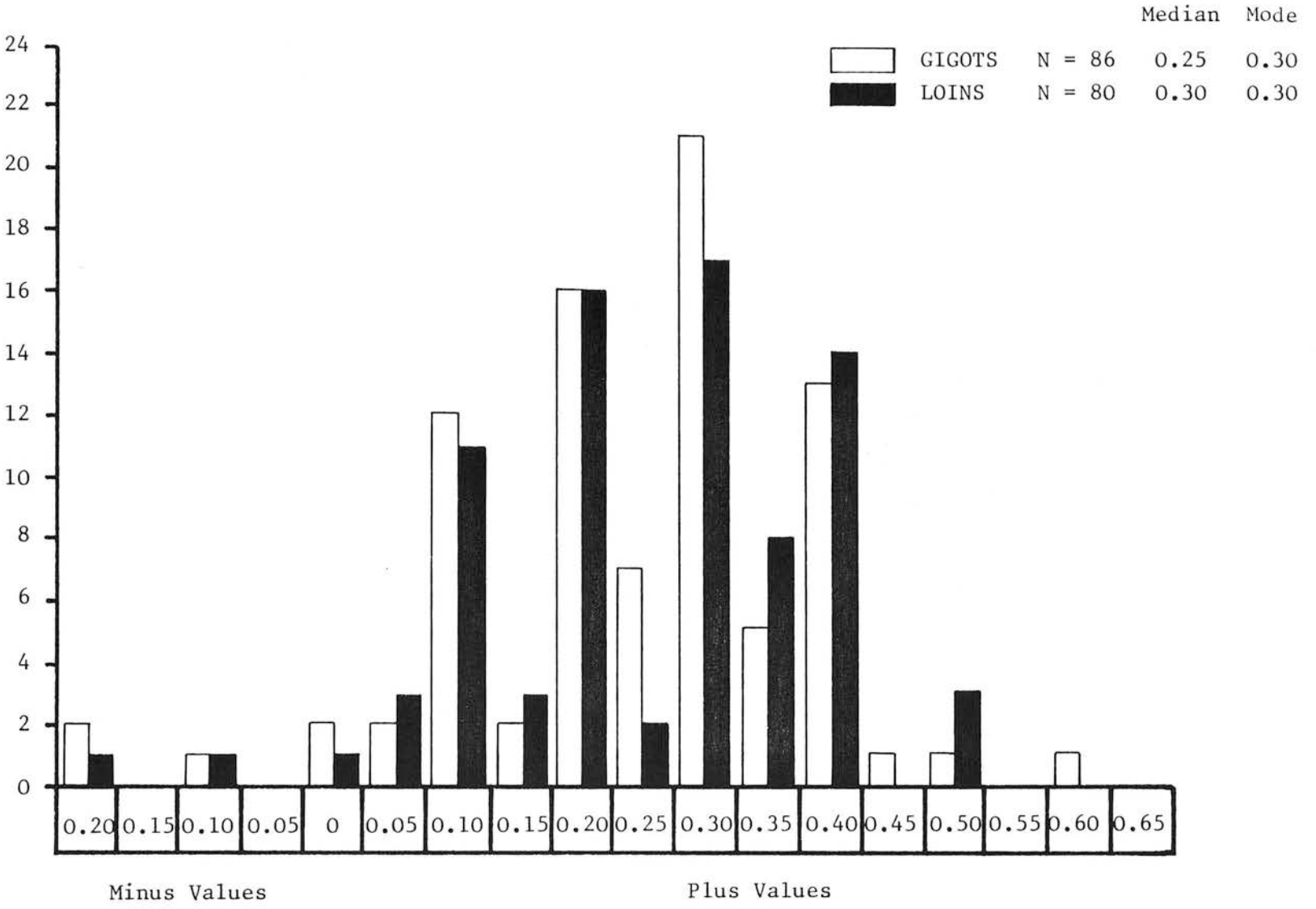

Figure 10.8 Frequency Distributions of Alteration in pH during the Cooking Process: Gigots and Loins

10.9 and 10.10. Although no statistically significant differences had been established between gigots and loins in the previous section and no differences existed between left and right sides, it was considered appropriate to combine gigots and loins from left and right sides for further analysis.

Table 10.16 Alterations in pH During the Cooking Process - Left and Right Gigots and Loins : Frequency Distribution

pH Values	Minus	LeG	RG	LeL	RL
0.20	"		2		2
0.15	"		0		0
0.10	"		2		0
0.05	"		0		1
0.00	No change		1		0
0.05	Plus	1	1	2	2
0.10	"	4	6	4	5
0.15	"	0	1	1	1
0.20	"	5	5	7	3
0.25	"	6	2	2	0
0.30	"	15	6	6	4
0.35	"	1	3	2	4
0.40	"	5	8	6	7
0.45	"	1	0	1	0
0.50	"		0		3
0.55	"		0		
0.60	"		1		
Mean		0.30	0.25	0.25	0.30
Mode		0.30	0.40	0.20	0.40
N		38	38	31	31

(Abbreviations are as in Table 10.1)

The Combined Median for the left and right gigots was 0.30 and the calculated value of χ^2 1.75 (1df) $p<0.20$. Corresponding values for left and right loins were 0.30 and 1.13 respectively ($p<0.30$).

Thus whilst no differences of statistical significance have been/

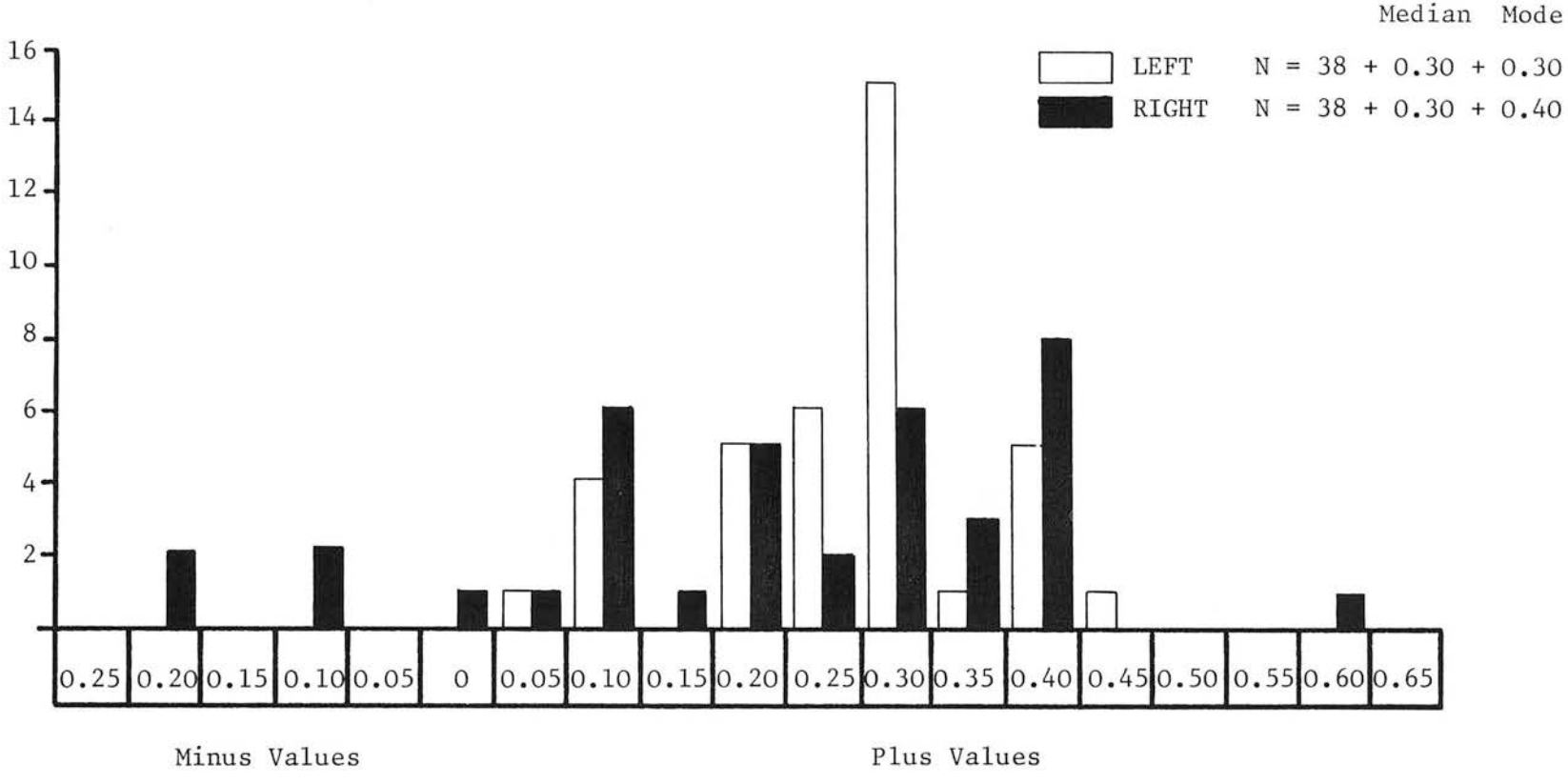

Figure 10.9 Alterations in pH during the Cooking Process - L. v R. Gigots (Irrespective of Initial pH Values)

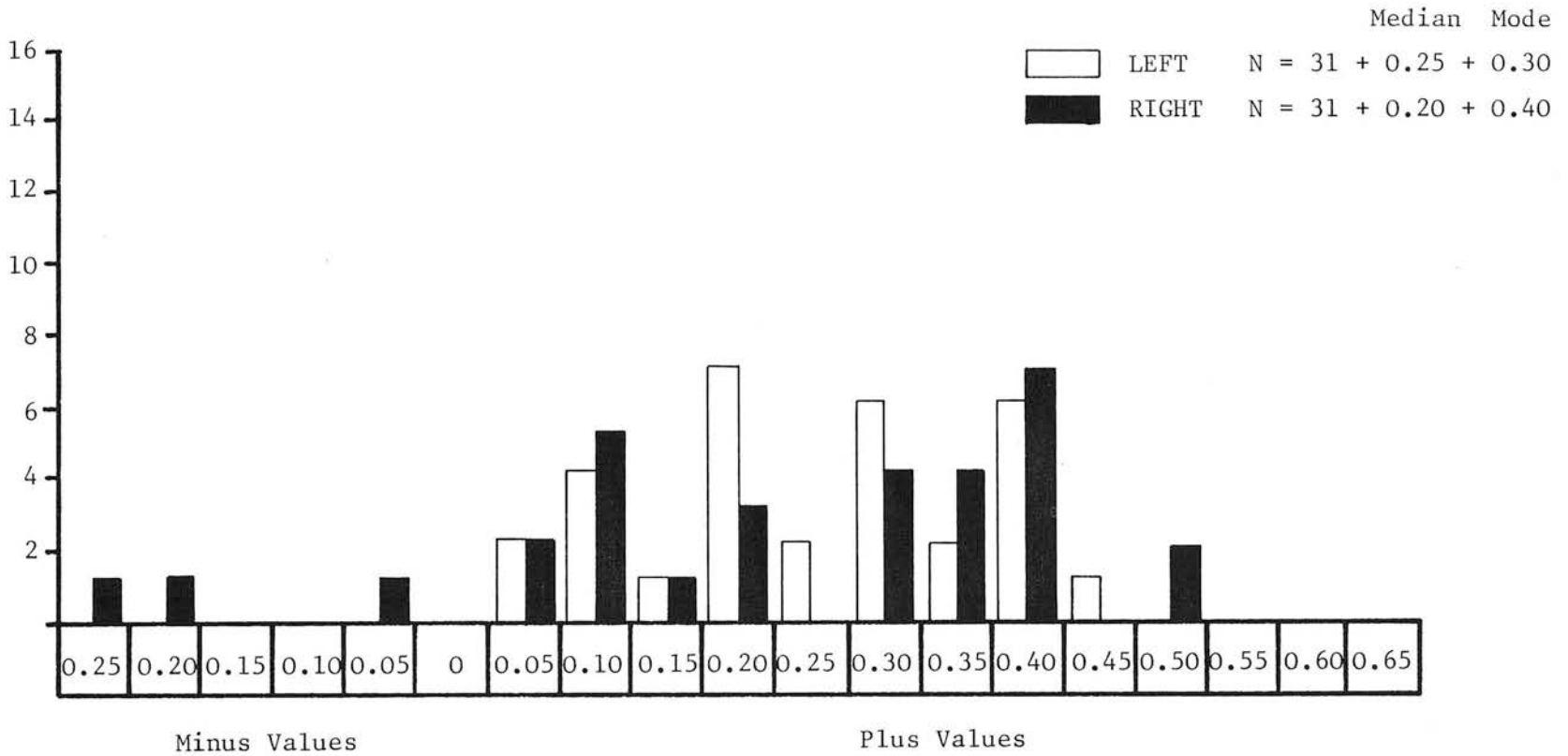

Figure 10.10 Alterations in pH during the Cooking Process - L. v R. Loins (Irrespective of Initial pH Values)

been established between left and right gigots and loins, changes in response to the cooking process are more uniform in left gigots and loins. Whereas pH changes in these joints are restricted to nine categories, corresponding categories for right gigots and loins are 17 and 15 respectively. Thus the response of the latter to the cooking process as reflected by alteration in pH value is more varied.

(v) Net Changes in pH Left versus Right Gigots and Loins

For the method by which the net changes were calculated the reader is referred to the closing paragraphs of Section 3 of this chapter. The calculated values are indicated in Table 10.17 and in Figure 10.11.

Table 10.17 Net Changes in pH in Left versus Right Gigots and Loins : Frequency Distribution

pH Values		Gigots	Loins
0.40	"	0	1
0.35	"	0	0
0.30	"	0	2
0.25	"	1	0
0.20	"	2	1
0.15	"	0	1
0.10	"	8	4
0.05	"	2	3
0.00	No change	9	7
0.05	Plus	2	4
0.10	"	3	1
0.15	"	1	0
0.20	"	3	1
0.25	"	2	2
0.30	"	1	2
0.35	"	3	1
0.40	"	0	1
0.45	"	0	0
0.50	"	0	0
0.55/			

Table 10.17 (Contd)

pH Values	Plus	Gigots	Loins
0.55	"	0	0
0.60	"	1	0
Median		0	0
Mode		0	0

In the Median Test, the Combined Median was 0 and the calculated value of χ^2 was 0.243 (1df) $p<0.70$. There is thus no statistically significant difference as indicated by the Median Test.

In view of the results of comparing left and right gigots and loins this result is not unexpected, since there are no statistically significant differences between raw and cooked pH values of left and right gigots and loins. Results of these comparisons are tabulated below. The full details are not tabulated here.

Table 10.18 pH Values of Left and Right Gigots and Loins i.e. Left versus Right*

	Gigots	Loins
N	38	31
Raw - Medians	5.75 5.825	5.5 5.5
Raw - Modes	5.90 5.85	5.5 5.5
Combined Median	5.75	5.5
χ^2	0.22 (NS)	0.364 (NS)
Cooked Medians	6.075 5.925	5.7 5.75
Cooked Modes	6.00 5.7	5.7 5.6
Combined Median	6.0	5.725
χ^2	0.05 (NS)	1.096 (NS)

*Left side value quoted first.

There are thus no statistically significant pH differences between left and right sides of either raw or cooked gigots and loins. The former, in both cases, show greater variation in pH.

Discussion/

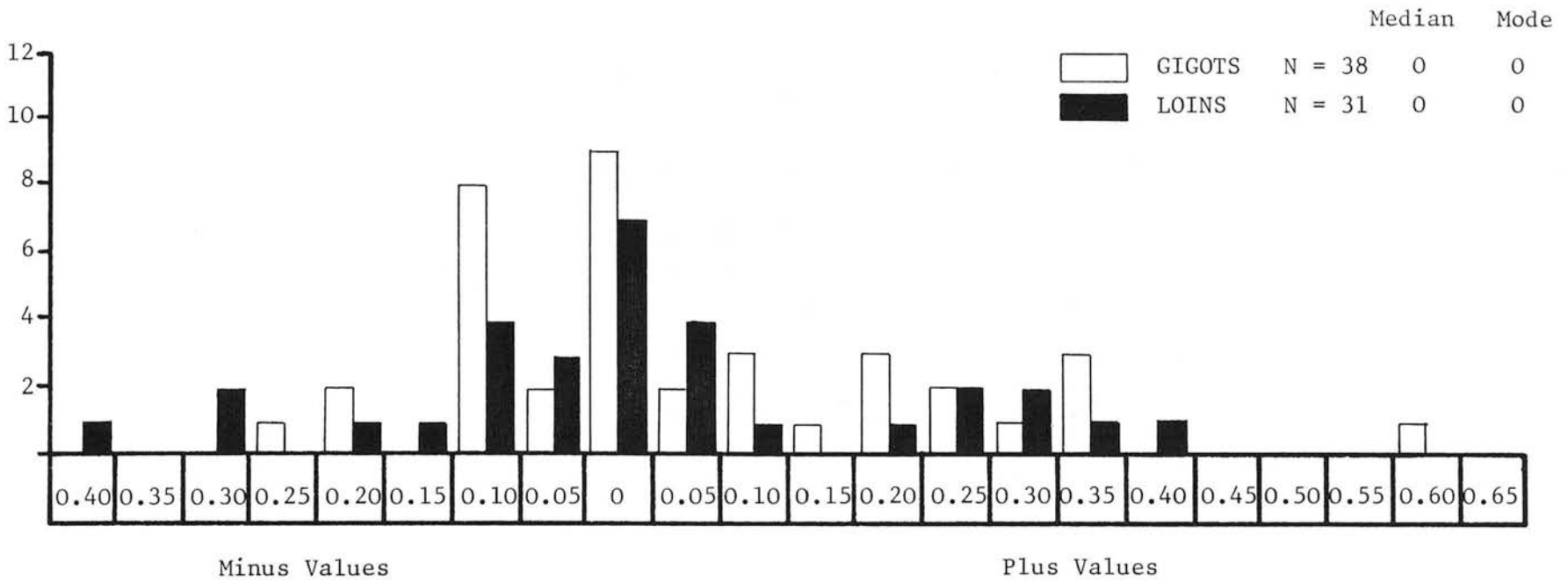

Figure 10.11 Net Changes in pH - Left and Right Sides Compared

8. Discussion and Conclusions

There are differences in initial weight of joints from left and right gigots and loins of standard anatomical location with maximum values of 70 - 74g. These differences are not of statistical significance.

There are statistically significant differences in total and evaporative weight losses between gigots and loins. These are lower in loins in both cases. This emphasises the importance of standardising the type of meat used for experiments. Variations within each category require large sample sizes to be available. There is no statistically significant difference between left and right gigots and loins in their reaction to the cooking process. This is reassuring, in that whilst either left or right joints were compared during many of the trials, occasionally it was necessary to compare joints from left and right sides of carcasses from two different feeding regimes.

In studying pH values, statistically significant differences have been demonstrated between pH values of gigots and loins both raw and cooked. Loins have lower pH values and tend to be more uniform. This applies to both the total sample of gigots and loins and to the 76 gigots and 62 loins which were used in comparing left and right sides of standard anatomical location. This difference in raw and cooked pH values does not affect changes in pH during the cooking process. There is no statistically significant difference in the response of gigots and loins to the cooking process as reflected by alteration in pH ($p<0.30$).

In comparing left and right sides of standard anatomical location, there are no differences in raw or cooked pH values, changes during the cooking process or net change between left and right sides.

Right/

Right sides have a tendency to show greater variation in pH than left.

In conclusion, loins are inherently less variable than gigots and may therefore be better suited for experimental purposes. It must however be borne in mind that their total and evaporative weight losses and raw and cooked pH values are significantly different from those of gigots.

CHAPTER 11

Summary

Subjects' performance in triangle tests on raw and cooked samples was compared. Results suggest that flavour differences caused by feeding regime could be identified by the use of small sections of the raw carcass prior to further investigations. Results of trials to assess the importance of fatty tissue in the aroma of beef and lamb and of pork samples suggests that further experiments would be desirable.

Comparing Volatiles in Raw and Cooked Lamb Samples and a Study of the Contribution of Fat to Meat Flavour

1. Following the deoiled herring silage trials with pigs, where differences in aroma of raw and subcutaneous fat were demonstrated between test and control samples, it was decided to compare the aroma of raw and cooked fatty tissue in the present lamb trials. Many substances likely to produce 'off' flavours are likely to be deposited in fatty tissue. In the trial with pigs, compounds causing aroma differences in raw samples either decomposed or volatilised during the cooking process since no preference was established for either test or control samples roasted under standard conditions. Although lambs are ruminant animals, grazing and forage crops as well as protected lipid supplements had been reported as causing 'off' flavours at the time the design of the present study was devised.

As indicated in Chapter 3, Wilcoxon's Matched-Pairs Signed-Ranks tests were used to compare subjects' performance in the three triangle tests (AAB, BAA and ABA) on raw and cooked samples which included lean and fatty tissue. The design of the experiment is given in Table 6.2 (Chapter 6) and the principle of the statistical technique/

technique is indicated in Chapter 3. In addition to trials carried out with special relevance to this study, experiments were carried out with BSc students of Agriculture at ESCA to determine the significance of fatty tissue in meat aroma in relation to lambs, comparing beef and lamb and pork and boar meats. Results of the Wilcoxon tests are given in context, but for clarity are also summarised in this chapter.

The scores of each subject for the three raw and the three cooked samples were compared. The test was one-tailed in that the hypothesis was that it would be easier to detect differences in the aroma of raw than of cooked samples. Results are as follows (Table 11.1). Calculated values of T which are equal to or less than values of T in Table J (Haber and Runyon 1977) where T denotes the smaller of the like signed ranks and DS is the total number of values with a different sign indicate that there is a statistically significant difference between scores.

Table 11.1 Wilcoxon's Matched-Pairs Signed-Ranks Tests - the Aroma of Raw and Cooked Samples

Grass v Rape Series 1 (GS1) (Chapter 6)

File	N	DS	T	Result
TA A	17	14	51.5	NS
B	18	13	44.5	"
C	15	10	16.5	"
D	13	9	31.5	"
E	17	12	25.5	"
F	13	7	7.0	"

Barley/

Barley v Turnip (BT) (Chapter 7)

	File	N	DS	T	Result
TA	G	15	11	20.5	NS
	H	20	15	47.5	"
	I	16	10	14.5	"
	J	8	6	5.0	"
	K	6	5	2.5	"
	L	6	4	3.0	"

Barley v Swede (BS) (Chapter 9)

T3	A	22	17	44.0	NS
	F	10	5	2.0	"
	L	6	5	2.5	"
	N	7	6	6.0	"

Barley v Cabbage (BC) (Chapter 9)

T3	H	22	16	68.0	NS
	J	17	13	45.5	"
	K	9	4	4.0	"
	M	5	4	4.0	"

Cabbage v Swede (CS) (Chapter 9)

T3	D	18	9	20.0	NS
	E	18	7	3.5	"
	I	19	14	30.0	"

Stubble Turnips (ST) (Chapter 9)

T3	B	23	20	47.5	$p<0.025$
	C	23	17	59.5	NS
	G	23	15	22	$p<0.025$

Thus in the few trials where statistically significant differences in the aroma of the A and B samples were detected, it was no easier to detect them before cooking lamb joints than after. The unexpected results in the Stubble Turnips trials have already been discussed in Chapter 9. It should however be noted that in relation to these two results of statistical significance in one trial the smaller sum of

like/

like signed ranks was positive and in the other negative. It thus seems reasonable to reject the hypothesis under test and to conclude that for samples of lamb under the conditions of the present trials, differences are not more readily demonstrated before than after cooking.

2. BSc students of Agriculture at ESCA have also carried out triangle tests as described in Table 6.2 (Chapter 6). In 1976 and 1977 they confirmed that differences in the aroma of subcutaneous fat existed between some of the pigs fed deoiled herring silage when they were compared with matched controls. Thus their results were in agreement with those obtained at QMC.

Experiments scheduled for November 1978 were deferred until January 1979. During the five sessions of experiments, interest was focussed not only on the aroma of lamb samples from different feeding regimes but also on the contribution of fatty tissue to species identification and in identifying differences between traditional pork and boar meats. The possible role of fatty tissue in meat flavour was considered in Chapter 2. The studies described were designed to assist in assessing its contribution to the aroma and flavour of meats. Results are presented in relation to the nature of the investigations rather than in chronological order.

In 1979, as indicated in Chapter 5, ESCA students carried out paired comparisons (preference) on the young barley fed lamb and the older turnip fed wethers. The latter were preferred ($p<0.01$). These findings were in line with those of other ESCA and QMC tasters.

3. In 1981, the ESCA students compared the aroma of roasted lamb from grass silage and lucerne silage fed lambs. The first, fourth and sixth series of tubes contained lean only. The second, third and fifth series contained lean + fat. Results are indicated in Table 11.2./

11.2. In the triangle tests, the lucerne silage fed sample was the odd sample with grass silage fed samples as the identical pair.

Table 11.2 Assessing Differences Between Grass Silage and Lucerne Silage Fed Samples by the Use of Triangle Tests

	Testers	Presentations	Correct Identifications	% Correct
Lean only	49	147	49	$33^{1}/3$
Lean + Fat	49	147	49	$33^{1}/3$

Each taster could identify the odd sample correctly on zero, one, two or three occasions. On the null hypothesis, expected frequencies are binomially distributed. If the observed and expected frequencies are compared χ^2 values are 4.79 and 3.03 for lean and lean + fat respectively (3df). As would have been expected from Table 11.2, there was no statistically significant difference between the aroma of the grass silage and lucerne silage fed samples. It was also clearly indicated that the presence of visible fat did not affect this result. The same group of students could detect no flavour difference by the use of triangle tests ($p<0.17$ derived from the z statistic) and had no statistically significant preference ($\chi^2 = 1.56$ (1df) $p<0.212$) for the lean of samples from either feeding regime.

As reported in Chapter 8, Table 8.10 statistically significant differences in the aroma of grass and rape fed samples were demonstrated. Values of χ^2 for raw and cooked samples were 11.93 and 3.85 respectively* The Wilcoxon test indicated that there was no difference in subjects' performance when samples were assessed raw rather than cooked. This ESCA result confirms results of previous experiments, including those results set out in Table 11.1.

4. In Chapter 2, it was noted that many workers consider that lean tissue from beef, lamb and pork is very similar when cooked identically

or/

*As noted in context, these results should be viewed with some caution. Probabilities are $p<0.001$ and $p<0.05$ (1df) respectively.

or water soluble extracts are heated. They suggest that the species specific factors are to be found in the fatty tissue.

An investigation of this hypothesis was carried out by ESCA students in 1980 by the use of triangle tests tabulated in Table 11.3. The design of the experiment is indicated in Table 6.2.

Table 11.3 Identifying Cooked Beef and Lamb With or Without Fatty Tissue by the Use of Triangle Tests

Beakers	1	4	6	A	Lamb + lamb fat
				B	Beef + beef fat
"	2	3	5	A	Lamb - lean only
				B	Beef - lean only

Presentation in each trial was ABA, AAB and BAA. Results are presented in Table 11.4.

Table 11.4 Results of Triangle Tests to Identify Beef and Lamb With and Without Fatty Tissue

	Tasters	Presentations	Correct Identifications	% Correct
Lean + Fat	54	162	93	57
Lean only	54	162	106	75

If observed and expected frequencies are compared χ^2 values are 81.94 and 145.92 respectively (3df). There are thus statistically significant differences in the aroma of beef and lamb in the presence or absence of visible fat. These findings, which were unexpected, are not in agreement with those of Wasserman and Talley (1968), Forss (1963), Hornstein and Crowe (1963a and 1963b), Pepper and Pearson (1971) and Pearson et al (1973).

ESCA students were not only able to distinguish beef and lamb samples in the absence of visible fat, but acuity was greater when only/

only lean meat of the species was used. When the results of the two trials were compared by constructing a 2 x 2 contingency table, the calculated value of χ^2 was 3.57 (1df $p<0.10$). Perhaps there was sufficient of the water soluble phospholipid fraction in the perimysium to confer typical flavour of species on the lean tissue during the cooking process (Rhodes 1973, Rhodes 1978, Mottram and Edwards 1983).

Unlike the triangle tests on flavour, there was no demonstrable difference in the appearance of the beef and lamb samples. Since only sniffing was required in aroma assessments, possible textural differences were not of relevance in these studies.

This group of students also carried out triangle tests on the flavour of the same beef and lamb samples. Although thin slices of the lean had been cut in uniform $10mm^2$ pieces, it is possible that appearance could have provided clues. By cooking the lamb to an internal temperature of 75^oC and the beef to 74^oC, doneness and hence colour was similar especially at the point of presentation. Textural differences, although every effort was made to minimise them, could have prompted a particular response. As in the triangle tests on aroma, there were two lamb and one beef sample presented. Subjects were however asked to identify differences only on the basis of flavour.

There were 63 tasters. Of the 63, 42 (67%) made the correct identification of the odd sample. The number of correct responses to give a statistically significant result is 34 ($p<0.001$).* Hence this result was clearly of statistical significance. With the qualifications of the previous paragraph, it thus seems that this group can differentiate cooked lamb and beef both on the basis of aroma and flavour. The presence of visible fat is not required.

This/

*Expanded Statistical Tables Roessler et al 1978.

This 1980 group of students performed well in all sensory appraisal tests they carried out. As indicated in Chapter 9, they established a statistically significant flavour preference for grass fed rather than grass silage with barley supplement fed lamb ($p<0.008$). Lambs were approximately the same age at slaughter and differed only in feeding regime.

5. In 1979, a group of BSc students of Agriculture undertook a series of sensory appraisal tests. Triangle tests followed the design indicated in Table 6.2. The paired samples (A) were young pork fat and the odd sample (B) was very mature pork fat. Samples were presented raw and cooked. Pork loins were used and cooked samples camefrom joints which had reached an internal temperature of 85°C. Results are given in Table 11.5.

Table 11.5 Results of Triangle Tests to Compare the Aroma of Raw and Cooked Fats from Young and Mature Pigs

	Testers	Presentations	Correct Identifications	% Correct
Raw	56	168	72	42.9
Cooked	56	168	74	44.0

There was a statistically significant difference in aroma of both raw and cooked samples ($p<0.01$). These differences were not altered by the cooking process. There was also a statistically significant difference in the flavour of the lean tissue as judged by triangle tests where 28 of 56 tasters identified the odd sample correctly ($p<0.01$).

This group also had a statistically significant flavour preference for turnip fed lamb slaughtered in February 1978 compared with young barley fed lamb slaughtered in May 1978 at five months of age ($p<0.01$) as indicated in Chapter 7.

In 1983, the aroma of roasted pork was compared with roasted boar meat./

meat. Both were loin roasts cooked to an internal temperature of 85°C. Interest was focussed on the importance of fatty tissue in pork flavour. There were 37 ESCA students of whom 23 were female and 14 male. Sample pairs (A) were pork and the odd sample (B) was boar. Tubes in the first, fourth and sixth trials contained lean + fatty tissue. Tubes in the second, third and fifth trials contained lean only. Results are given in Table 11.6.

Table 11.6 Results of Triangle Tests to Compare the Aroma of Roasted Pork and Boar Meats With and Without Visible Fat

	Testers	Presentations	Correct Identifications	% Correct
Lean + Fat	37	111	85	76.6
Lean only	37	111	49	44.1

Thus, unlike the results of the species identification trials (Table 11.4) the presence of fatty tissue greatly improved the testers ability to detect the odd sample in triangle tests. A 2 x 2 Contingency table was constructed to demonstrate the distribution of correct and incorrect responses. Observed and expected frequencies were compared and the value of χ^2 was calculated. This value was 23.06 (1df) indicating that the presence of fat made a very highly significant contribution to subjects' performance. The corresponding value of χ^2 in the species identification, where performance was better in the absence of fat was 3.57 (1df) $p<0.10$.

Compounds responsible for sex odour (formerly known as boar taint) are ether soluble, non-saponifiable and located in fatty tissue. Two compounds have been identified - 5 -androst-16-ene-3-one and 3 -hydroxy-5 -androst-16-ene of which the 3-ketone is considered to have the more important role (Reineccius 1979). There is a sex difference in ability to/

to detect sex odour. Griffiths and Patterson (1970) reported that 46% of males and 98% of females could detect the aroma. Thus the performance of the males and females in the tests was studied. Results are indicated in Table 11.7.

Table 11.7 Comparison of Performance of Males and Females in Triangle Tests when Roasted Pork and Boar are Assessed With or Without Fat

	Lean + Fat			Lean Only		
	Correct	Incorrect	Total	Correct	Incorrect	Total
Males	30	12	42	12	30	42
Females	55	14	69	37	32	69
Totals	85	26	111	49	62	111

Values of χ^2 were calculated for the two trials and were 0.74 and 5.63 for lean + fat and lean only respectively (1df). Thus in these trials male and females identified the odd sample in the triangle test in approximately the same proportions (71.4 and 79.7% respectively). Whereas if only the lean were presented, where the value of χ^2 was 5.63 (1df) there is a statistically significant difference ($p<0.02$) between the responses of males and females with correct responses of 28.6% and 79.03% respectively. Thus greater sensitivity in females is confirmed. Sufficient fat is likely to be present in the lean tissue to identify the odd sample presumably as a result of the presence of these sex odour compounds in the perimysial tissue.

If the total correct responses in the two trials is considered, Table 11.8, the probability of achieving these results by chance alone is indicated. The Expanded Tables of Roessler et al (1978) were used to determine these values.

Table 11.8/

Table 11.8 Responses in Triangle Tests when Roasted Park and Boar are Assessed With and Without Fat

	Lean + Fat			Lean Only		
	Total	Correct	p<	Total	Correct	p<
Males	42	30	0.001	42	12	NS
Females	69	55	0.001	69	49	0.001

Whilst the variation in response between males and females suggests that sex odour may be the cause of the detection of difference in the aroma of pork and boar meat, without further study this cannot be confirmed. It is also recognised that in the presence of visible fat, performance of males and females in triangle tests was comparable.

This student group also carried out triangle tests using the lean of the pork and boar meats. There was no detectable difference in flavour as is indicated in Table 11.9.

Table 11.9 Comparing the Flavour of Pork and Boar Roasts by the Use of Triangle Tests

Tasters	Responses	Correct Responses	% Correct	Significance
Males	13	5	38.5	NS
Females	21	9	42.8	NS
All Subjects	34	14	41.2	NS

Thus the different response of males and females to the aroma of the lean boar meat is not reflected in the triangle tests on flavour. However, Reineccuis (1979) emphasises that higher temperatures may be required to allow compounds such as those responsible for sex odour to volatilise. Samples for aroma assessment were held in a waterbath at 60°C whereas flavour was assessed using cold meats. Temperature difference could thus account for the disparity in results of the two tests. Similarly, there was no statistically significant preference for/

for either sample when lean portions were tasted cold (Table 11.10).

Table 11.10 Paired Comparison (Preference) Tests - Pork and Boar Roasts

Tasters	Responses	Boar Preferred	% Preference	Significance
Male	15	9	60.0	NS
Female	23	8	34.8	NS
All Subjects	38	17	44.7	NS

Statistical significance of the results presented in Tables 11.9 and 11.10 was determined by the use of the Expanded Statistical Tables (Roessler et al 1978). From Table 11.10 it will be noted that fewer females than males preferred the boar sample. This preference was not of statistical significance ($\chi^2 = 1.00$).

6. In conclusion, in the experiments described, it has been shown that it is no easier to detect differences in aroma arising as a result of varying feeding regimes in raw than in cooked lamb samples. This conclusion was reached both by comparing judges' performance in triangle tests using raw and cooked samples and by comparing scores of each judge for raw and cooked samples by Wilcoxon's Matched-Pairs Signed-Ranks tests.

Contrary to suggestions by other authors, the presence of visible fat was not necessary to enable detectable differences in the aroma of roasted beef and lamb samples to be demonstrated. Although there may have been 'prompts' which were not present in the aroma trials, detectable differences in the flavour of the lean meats were also demonstrated. Further experiments are required to confirm or refute these findings.

Statistically/

Statistically significance differences were demonstrated in the aroma of fats from traditional pork and joints from very mature pigs. In this experiment, unlike the deoiled herring silage trials, there was no difference in judges' performance when fats were assessed raw or cooked. Statistically significant differences in the flavour of the cooked lean samples was demonstrated. Thus neither the cooking process nor the absence of visible fat affected judges ability to differentiate aroma or flavour.

When the aroma of cooked pork and boar meats was compared, the presence of visible fats made a very highly significant improvement in judges' performance. In comparing beef and lamb judges' performance in triangle tests, acuity was greater when assessments were made in the absence of visible fat. This greater acuity was not of statistical significance ($p<0.10$) whereas in the pork and boar trials the improvement was ($p<0.001$).

In comparing cooked pork and boar meats, performance of female and male judges in the aroma trials was compared. Performance of females was comparable either in the presence or absence of visible fats. Performance of male judges deteriorated in the absence of visible fats and statistically significant differences in pork and boar aroma were not demonstrated. This finding suggests that compounds responsible for sex odour caused detectable differences in aroma between samples. No flavour differences were established in triangle tests and there was no statistically significant preference for either sample, although females tended to prefer pork rather than boar. It should be noted that aroma assessments were made at approximately 60^{o}C whereas tasting took place at room temperature. Volatiles responsible for differences in aroma at the higher temperature may not be/

be detectable at room temperature. Further tests would have been required to demonstrate if compounds responsible for sex odour, which are known to be volatile only at higher temperatures and to cause different responses in female and male subjects, caused the detectable aroma differences. Cain (1982) reported general female superiority in aroma tests with 46 subjects.

It is also suggested that the role of fats in the aroma and flavour of meats cooked by traditional methods requires further investigation. It seems likely that costs could be minimised by the use of triangle tests on raw lamb tissue to detect possible differences in flavour arising as a result of feeding regime. Tests could be readily organised and the value of carcasses preserved. The only problem could be the disagreeable effects the procedure has on so many judges.

CHAPTER 12

Summary

The problems of comparing results from different laboratories are considered. Variations in the preparation and presentation of samples to tasters, in assessment procedures and in the analysis of experimental results are described and discussed. Results of the present studies are considered in detail. Recommendations for further investigations are made.

1. Problems of Comparing Results Between Laboratories

As has already been noted, in reviewing studies of lamb palatability, the reader is immediately aware of striking differences in the preparation, presentation and methods of assessment between laboratories. The majority of publications indicate that some form of hedonic rating scale has been used to assess lamb characteristics. Scores derived from these scales are used in subsequent statistical analysis of data. Characteristics studied include intensity of flavour and aroma on a 12 point scale (Cramer et al 1967; Nicol and Jagusch, Experiment 3 1971; Misock et al 1972), a nine point attitudinal (like/dislike) scale, the development of which was traced by Peryam and Pilgrim (1957), (Kemp et al 1972 and 1976; Jeremiah et al 1972) or more structured scales such as those described by Woodhams et al (1965), Kirton and Pickering (1967) Park et al (1972), and Harries et al (1963). The latter authors devised their vocabulary after taking into account tasters' views concerning ease in use and clarity of meaning. Their scales involve the use of positive and negative integers whereas those of Park et al (1972) do not. The use of hedonic scales of different types is considered in Chapter 3 of this study. Harries (1963) indicated that it was recognised that the scales were arbitrary and that it was not known by how much steps between successive/

successive scores may vary. Detailed comparisons of the effectiveness of different scales was not possible at that time (1963). At MRI, a combination of an attitudinal scale and a more structured scale was used by Rhodes (1976).

Most of the results of sensory testing are studied by analysis of variance using the techniques of Snedecor and Cochran (1967). Arithmetic means and correlation coefficients are calculated and t tests are used. Amerine et al (1965) note that nine point attitudinal scales, first described by Peryan and Girardot (1952), were designed for use with observers entirely without experience in food testing such as the American soldiers used in their studies and for consumer studies. The advantages and disadvantages of hedonic rating scales are also indicated in Chapter 3 of the present study.

In Chapter 2, the difficulty of flavour assessments in complex food systems such as cooked meats is considered. For instance, Rhodes (1978) posed the queries "What constitutes delicious odour, how can it be maintained, enhanced, intensified and reproduced?" In the same paper he indicated that success in analysing desirable cooked meat flavour has not yet been achieved. Over a period of time and accepting their limitations, hedonic rating scales have been used at MRI. In an experiment to compare the eating quality of U.S. and U.K. meats cooked to internal temperatures typical of U.S. and U.K. practice, mean scores for flavour did not differ significantly between origin and cooking method, but significant panellist x origin and panellist x cooking method interactions were demonstrated. Eight of the 14 panellists found no flavour differences between the two samples of different origin. Of the remaining six panellists, two consistently identified flavour differences whereas the other four only did so when cooked/

cooked by one or other method. He acknowledged that the limitation of the attitudinal hedonic rating scale used was immediately apparent and that such results did not allow firm conclusions to be drawn about the flavour of the meats from either source. 'Meat flavour intensity' whilst more objective than hedonic rating scales can be of use in making direct comparisons but hedonic bias is still present in that panellists must decide individually what is meant by meat flavour before recording its intensity.

These observations made by such a highly experienced worker in the field of sensory appraisal of meats suggest that some of the findings previously reported should perhaps be regarded with a degree of caution particularly when differences between samples are slight. Rhodes implied that different approaches to sensory appraisal techniques in relation to meat flavour might be considered. Even in 1963, Harries (ibid) was aware of some of the constraints of using hedonic rating scales. These constraints will be examined later in this chapter.

As indicated in Chapter 1, the main aim of the present study was to determine if any of the fattening regimes for store lambs used in the 1977 ESCA project - and where appropriate the 1978 replicate - caused flavour defects in roasted gigots and loins. Rape fed lambs had been shown by Park et al (1972) and Wheeler et al (1974) to have objectionable aroma and flavour when minced samples were cooked. Lucerne had also been implicated in producing flavour defects (Nicol and Jagusch (1971), Park et al (1972a, 1972b and 1975). Protected lipid supplements have also been reported as causing differences in aroma and flavour in lamb although these differences were not always regarded as objectionable (Wright 1974; Garrett 1976).

In Chapter 2, causes of variation in lamb flavour are considered.

In/

In the live animal there are conflicting reports on the effects of breed, sex and age on lamb flavour although there is general agreement that flavour intensity is greater in meats from mature than from young animals. Since all three of these factors varies between laboratories, it is not perhaps surprising that different experimental findings are reported. Some reports of the effect of preslaughter diet on lamb flavour are indicated in Chapter 1. In addition, preslaughter stress introduces yet another variable affecting as it does muscle glycogen concentration and hence ultimate pH with the implications considered in Chapter 2.

Post slaughter, still more sources of variation operate. Conditioning not only produces meats of greater tenderness but causes changes which, when properly controlled, allow attractive flavour to develop during the cooking process. Carcass ripening practices will inevitably vary from one laboratory to another.

2. Preparation and Presentation of Samples to Tasters

In addition to factors described above, cooking procedures which exert such marked effects on flavour are not standardised between laboratories. Texture has both direct and indirect effects on flavour perception. Release of flavour volatiles and water soluble flavouring compounds will obviously be influenced by texture. Even if assessment of only flavour is required, it is difficult, if not impossible, to prevent texture exerting some influence on flavour perception. For this reason, many workers have used minced samples for presentation to tasters either as stews, meat loaves or patties in order to minimise any distracting effects of texture. Wong et al (1975) combined 90 per cent minced lean tissue with 10 per cent fat, added/

added half the combined volume of water and a 'little' salt. This mixture was held at 150°C for 30 minutes before 5g samples were presented to tasters for assessment of intensity of mutton flavour. In this study the chemistry of mutton flavour was under investigation particularly the contributions of 4-methyloctanoic (hircinoic) and 4-methylnonanoic acids to mutton flavour. Subjects were sensitive to concentrations of 0.5ppm when these compounds were added to bland meat. Batcher et al (1962) and Jacobs et al (1972) used lamb patties in comparing the flavour scores of meats from rams and wethers. No flavour score difference was present between control and treated groups. Park et al (1972) found meat loaves less sensitive than stewed minced lamb although Wasserman and Talley (1968) had found the former method of presentation to be satisfactory. Park et al (1972) based their judgement on a series of ten triangle tests when flavour intensity differed. Discrimination was significantly poorer ($p<0.05$) with baked loaves than with stewed minces. In addition it was noted that minces were more readily cooked in a standard manner. Trimmed lean together with added separated fat (10%) was minced. A batch of 400g without added salt was brought to the boil with 400g of water in a covered pan. After 20 minutes samples were held in a water bath at 75°C before presentation (hot) to the assessors. Pearson et al (1973) also used defatted lean minced tissue but this treatment is not typical of normal culinary practice particularly as the aim of Pearson's studies was to investigate the contribution of fat and lean to the aroma of cooked beef and lamb. Flavour assessments were not made. These workers however suggested that their lamb samplesdid not contain the strong aroma associated with lamb grazed on white clover and lucerne as reported by Cramer et al (1967) and Nicol and Jacgusch (1971).

In/

In later trials described by Ford and Park (1980), lamb carcasses suspended from the pelvis within one hour of slaughter (tenderstretched) were used. Carcasses of more uniform tenderness are produced. Thus "the risks of tenderness interfering with flavour assessment are therefore greatly reduced and meat can be presented to panellists in a more palatable form." The influence of cooking method on panel results has not been measured but in a series of trials carried out by Park et al (1975) results suggest that interference with flavour assessment did not occur even if samples were not presented as uniformly textured mince. The techniques of preparation of roasted samples which were used do not however seem to be as highly controlled as in some other laboratories. Rear legs were roasted uncovered on trivets on stainless steel trays in a convection oven at 190°C with a cooking time of six minutes for each 100g. Forelegs, prepared as rolled shoulders were roasted either uncovered or in polyester cooking bags to an internal temperature of 77°C. The principles of these two heating methods differ in that the uncovered shoulder was cooked by a dry and the covered shoulder by a moist method of cooking. Time/temperature relationships alter. Flavour differences are often observed when these two techniques are compared. Slices of 6-9mm thickness were either served hot within 30 minutes of cooking or cold after holding overnight in polyethylene bags at 2°C. These practices were followed at Cannon Hill but at North Ryde legs were roasted in polyester bags at 175°C for 125 or 150 minutes for lamb and mutton respectively. Comparison of performance in assessing mince with 'roasted' samples is thus not easy from the data which is presented although it is derived from the same experiment (Park et al 1975).

Most American workers have followed the cooking procedures indicated by/

by Cramer (1967). In the experiments described, untrimmed chops from the 12th rib were cooked on stainless steel racks in individual covered casserole dishes. No water was added. A preheated oven (155°C) was used. Chops were cooked to an internal temperature of 74°C. After cooking, lean and fat were separated into $10mm^2$ cubes. Residual meat and bones of each chop were placed in wide-mouthed stoppered glass jars for odour evaluation. Tasters were required only to score flavour intensity and to ignore other organoleptic factors. After lean samples had been assessed, fats were subsequently chewed and assessed. Following flavour evaluation, the contents of the jars were assessed for aroma intensity.

Nicol and Jagusch (1971) based their work with trained panellists on that of Cramer (1967). From the description of the procedures in their third experiment procedures appear to be the same as those of Cramer. Carpenter et al (1969 and 1970) roasted lamb joints uncovered at an oven temperature of 177°C to an internal temperature of 75°C. It should be noted that this was the procedure followed in the present study. Usborne (1961) roasted covered leg roasts to an internal temperature of 180°F (82°C). Batcher et al (1962) using an oven temperature of 163°C also allowed joints to reach this internal temperature (82°C). Oliver et al (1967) allowed twelfth rib chops to reach an internal temperature of 70°C when grilled. Smith et al (1970) are likely to have used the same cooking procedures as judged from a report of their experimental studies. Woodhams et al (1965) and Kirton et al (1967) used an oven temperature of 163°C to achieve internal temperatures of 80°C. At MRI, Bristol, oven and internal temperatures are reported as 150°C and 80°C respectively for meats prepared for sensory appraisal. Examples are quoted in Harries (1960 and/

and 1963), Rhodes (1971 and 1976) and Dransfield (1979).

Thus cooking procedures and hence time/temperature relationships show considerable variation. This variation is likely to affect some of the changes in meat induced by the cooking process. Certainly there are considerable flavour differences between minced samples served as stews and roasted meats. Studies of the differences in roasting techniques which are extensively reported in Griswold (1962), Paul and Palmer (1975) and Campbell et al (1979) concentrate more on the effects of oven and internal temperatures on weight losses, tenderness and juiciness than on flavour. In Chapter 5, tasters found considerable difficulty in differentiating meats cooked to different internal temperatures. There may thus have been little difference in flavour particularly between those cooked to internal temperatures of 75°C and 80°C. It may be that, particularly for lamb, differences in roasting procedures do not produce sufficient difference in flavour to make comparisons of results from different laboratories a problem.

3. Variations in Assessment Procedures

As indicated earlier in this chapter, additional variations occur when findings of different laboratories are compared both because experimenters vary in how assessments of cooked meat samples are to be made and how experimental results are subsequently analysed. For instance, Cramer (1967) required only intensity of flavour to be scored in one of five categories. These categories were 'none', 'slight', 'moderate', 'strong' or 'intense'. Within each category except the first (scored zero) panel members marked the score sheet with a minus sign, a tick or a plus sign. One point was then allocated/

allocated for each of these three responses thus producing a twelve point scale. Results were tested for significance by analysis of variance according to the methods of Snedecor (1956). Nicol and Jagusch (Experiment 3 1971) and Misock et al (1972) also followed this practice. Many experimenters including Kemp et al (1972 and 1976), Jeremiah et al (1972), Smith and Carpenter (1970), Headley and Jacobson (1960) and Kirton and Pickering (1967), required attributes such as tenderness, juiciness, flavour and general/ overall acceptability to be assigned scores on a nine point attitudinal (like/dislike) scale. The development of such attitudinal scales was traced by Peryam and Pilgrim (1957). This method, first described by Peryam and Girardot (1952) originated from a seven point scale devised to provide information on the preference of army personnel for items on their menus. This type of unstructured scale was used in the comparisons of lamb samples cooked to varying internal temperatures described in Chapter 5. Combined with possibly marginal flavour differences indicated earlier in this chapter, the use of such unstructured scales may have been daunting particularly to tasters who were relatively inexperienced. Both factors may have contributed to variation and lack of consistency in responses. Because of the importance of selection of optimal internal temperature for the tasters involved, considerable study of these results was made. Although an internal temperature of 75°C was eventually selected, studies in other laboratories suggest that internal temperatures between 75°C and 80°C are acceptable to most tasters.

Harries et al (1963) acknowledged that more structured hedonic rating scales may be required to make them easier to use by clarifying the meaning of each of the terms used to describe product attributes.

In/

In the Food Industry where a limited number of products is assessed as part of a quality control or quality assurance programme, this is recognised practice. The system described by Harries (1963) was being devised at the same time as Shewan et al (1958) were developing a scoring system for wet white fish. Here the object was to examine deterioration of quality with storage time. These workers were able to specify a sequence of odour and flavour descriptors related to age after catching. This allows fish to be classified on a ten point scale. As indicated by Howgate (1977) this system is still used. Odour and freshness assessment of breads is made on a six or eight point scale by the use of descriptions such as absolutely fresh through to absolutely stale (Cornford 1977).

It was indicated that only the descriptors are supplied to tasters. Scores are allocated after tasting sessions to avoid the problem highlighted by Shephard (1955) who suggested that if scores and descriptive terms are used simultaneously participants may rate primarily by scores thus using a linear scale. This particular scale resembles the structured hedonic rating scales used in the assessment of meats more closely than the Shewan scale for fish. In meats, the aim is to study the effects of production variables and not to study deterioration.

Misock et al (1972), in studying the flavour of wethers and rams slaughtered at 183, 237 and 295 days used a simple type of hedonic scale to measure flavour and aroma. This nine point scale ran from one representing no flavour or aroma to nine for characteristics described as very strong. The scale devised by Harries et al (1963) studied meat flavour, juiciness, tenderness, 'degree of doneness' and off flavours. Intensity of flavour was scored from zero to five with zero/

zero being tasteless/insipid and five strong. Juiciness was rated from very juicy to very dry on a ten point scale. Tenderness was rated in the same way. 'Degree of doneness' was estimated on an eleven point scale from 'grossly overdone' through 'neither overdone nor underdone' (zero) to 'grossly overdone'. 'Off' flavours were zero when none was present through five points to 'very strong 'off' flavour'. Tasters were required to specify the nature of 'off' flavour detected. Rhodes (1976) used an eight point scale to assess darkness/lightness, tenderness/toughness, juiciness/dryness and overall acceptable/unacceptable characteristics. Flavour was assessed on an eight point like/dislike scale. Batcher et al (1962) scored tenderness, juiciness, natural lamb flavour and off flavours on ten point scales. Very tender to very tough, very juicy to very dry, natural lamb flavour from none (one) to pronounced (ten) were the descriptors used. Unlike the scales of Harries and Rhodes described above, only positive integers were used. Headley and Jacobson (1960) used an eight point scale for tenderness, juiciness and flavour where zero was very poor flavour and seven very good. In two consumer trials, Nicol and Jagusch (1971) used a nine point scale where nine indicated a high degree of acceptable flavour, tenderness and juiciness. Woodhams et al (1965) used a nine point scale similar to that of Rhodes but only positive integers were used.

Park et al (1972b) asked tasters to identify samples from crop fed lambs in a duo-trio test comparing pasture grazed lambs with those fed winter forage crops. They were then asked to assess the intensity of aroma and flavour of both as normal or 'meat' and abnormal or 'foreign' flavour on a nine point scale, (zero for 'none', eight for 'very strong'). In addition, acceptability of the flavour of the two samples was assessed on a nine point scale where zero was 'nauseating' and eight 'very good'. These/

These workers point out that correct responses are largely a function of the difference between pasture and crop-fed meats whilst "answers to the questions which followed were useful in determining the size of differences and preferences if any." Additional comments to identify the nature of any taints or 'off' aroma or flavour were recorded on special score sheets which included a list of adjectives derived from Harper et al (1968a) and Harper et al (1968b). Descriptions were in five categories (Ford and Park 1980) and differed for aroma and flavour. As well as rancidity characteristics, 'foreign' or 'other' aromas and flavours could perhaps be described as taints arising as a result of diet, or postslaughter absorption of contaminants from the surrounding environment, sanitisers and packaging materials. The final descriptions for aroma and flavour include such attributes as 'musty', 'faecal', 'putrid' and 'sickly'.

Statistical analyses were made in the duo-trio tests by the use of appropriate tables. The proportion of correct decisions was used to determine the effects of preslaughter treatments. Analysis of variance was carried out on angular transformations of the ratio of correct to total responses in each test. The significance of the differences in flavour properties and treatment effects was determined by the application of the t test. Simple correlation coefficients using mean taste panel scores were calculated.

Park et al (1972a) were advised on experimental design by Dr A Howard of CSIRO Meat Research Laboratory and on some of the statistical treatments by Dr R Baxter of CSIRO Division of Mathematical Statistics, Brisbane. Their experiments are quite different to others which have been reviewed in the present study. The Duo-Trio is a well documented and recognised difference test which is considered highly suitable for identifying/

identifying differences between samples. Possible distracting effects of requiring additional assessments during difference testing are noted in Chapter 3. Ford and Park (1980) indicate that in many experimental studies only hedonic scoring systems are used to measure the palatability of meats. They consider that such results indicate only the personal preferences of a particular group of people and may not be meaningful unless a large panel has been used. Since most objections to Australian and New Zealand sheep meat flavour come from North America and Japan, they considered that it would have been desirable to use consumers from these countries as tasters. This was not however practicable. Their panels were experienced tasters trained to detect subtle differences in flavour who were more likely to be sensitive to aromas and flavours which the overseas consumer might find objectionable although the tasters themselves did not. The intensity of aroma and flavour assessments were intended to be as objective as possible thus and/should bear no relation to panellist's personal preferences. Increasing flavour acceptability was, as indicated, made on a nine point scale using descriptors from 'nauseating' to 'very good' and a score card was used to identify the nature of "off" aroma or flavour or taint.

Whilst fully appreciating the frustrations and limitations of the use of hedonic rating scales, the author is somewhat doubtful about the use of some of the terms used as descriptors even with experienced tasters. That the possibility of a sample being presented which could be described as nauseating, putrid or faecal unless tasters had previously been warned that meat taint was under investigation seems likely to introduce an element of bias into judgements. The probability of introducing Type II errors seems also likely to be increased.

At/

At the time of the present study, the author was a little uneasy about some aspects of the procedures indicated by Park et al (1972a) particularly those concerned with flavour acceptability and foreign aroma and flavour discussed earlier in this section. It was not until further details of their experiments were published (Ford and Park 1980) that reservations about findings that flavour defects in lamb may arise as a result of differences in feeding regime became greater.

Procedures were modified slightly in studies on the effects of a protected lipid supplement (Park et al 1975). On this occasion flavour acceptability was rated on a hedonic scale where zero was very poor and eight was very good. Taste panel members were also provided with a list of descriptors of odour and flavour qualities to assist in identifying the nature of any different aroma or flavour. No details of the descriptors was given but descriptions of aroma and flavour given by panel members did not include such emotive terms as nauseating, putrid, or faecal. No acknowledgement is given in this report of any assistance with experimental design. Analyses of variance in scores between treatments were made and where significant differences were established Duncan's multiple range tests were used to identify their source.

4. Variations in Analysing Results

Harper (1977), tracing the history of the sensory appraisal of foods, indicated that its present development is a post-1945 phenomenon which slowed down after 1952 except in a few research institutes and university departments. Current interest appears to be focussed on the use of existing tests with appropriate variations for different products, comparing the efficiency of existing tests and more recently on statistical analysis of experimental data. At symposia the first aspect normally receives/

receives the greatest emphasis. Some comparisons of the efficiency of different test procedures are quoted in Chapter 3 and, as indicated, comparisons of triangle and duo-trio tests are being made at present in the author's laboratory.

Von Sydow and Akesson (1977) make a comparison of parametric and nonparametric approaches to sensory data. They indicate that parametric analysis is based on the following assumptions:

(a) sensory attributes to be evaluated are perceived as a continuum with respect to magnitudes of intensity

(b) subjects are able to estimate with consistency the perceived magnitudes of intensity in numerical or equivalent scale categories

(c) the distribution of responses follows some parametric distribution such as normal or log-normal distribution.

Assumptions (a) and (b) imply that data is interval scaled.

The nonparametric approach involves less stringent assumptions:

(d) the sensory attributes may or may not be perceived as a sensory continuum

(e) subjects can only rank or classify stimuli into a set of discrete categories

(f) parametric specifications of the underlying distribution of responses are not made.

This nonparametric approach implies that sensory data are ordinal or nominal in nature. Despite resulting difficulties in some quantitative analyses, these authors believe that the nonparametric approach is justified.

Williams (1978) discusses the problems of handling data derived from hedonic rating scales. Arithmetic means may be justified when dealing with objective data, but their use is questionable for subjective data such/

such as hedonic scales. It is possible that a mean score reflects the opinion of no assessor or group of assessors. Results are unlikely to be as clear cut as the mean suggests. This author suggests caution when drawing conclusions from panel means.

As early as 1956, Siegel considered that the attitudinal scales used by the behavioural scientist were at best ordinal. Many experimenters who assess foods use such like/dislike attitudinal scales which are essentially similar to those used in psychometric studies. Harries (1963) admitted that it was not known by how much steps between successives may vary. Rhodes (1979) recognised the limitations of attitudinal scales.

At a meeting of the Royal Statistical Society (Edinburgh, March 1982) Smith read a paper under the title 'Statistical Methods and Problems in Food Tasting'. He indicated that linearity and normal distribution cannot be assumed and that the use of nonparametrical statistics for analysing data from hedonic rating scales was preferable. In subsequent correspondence he confirmed his concern over the free use of statistical tests based on the normal distribution without checks on their validity and highlighted the warning which appears in BS 5929 Part I (1980). His present recommendation is that normality should be checked by examination of histograms. Although the scale is not an equal interval scale, with a large panel, distribution may approach normality and the error of using parametric statistics may be small. He considers it unwise however to quote significance levels.

Piggott and Land (1981) acknowledged some of the difficulties in using parametric statistics. Howgate (1981), describing the use of nine point attitudinal scales where one corresponds to dislike extremely and nine to like extremely for the assessment of fish products, did not find/

find that responses were normally distributed. If median values were five, distribution was symmetrical but flattened. Distribution-free statistical procedures such as the Kolmogorov-Smirnov test or the Wilcoxon Matched-Pairs Signed-Ranks test were used to determine the significance of differences between products. Because of varying shapes among distributions - although the median is to be preferred as the measure of central tendency - a better index could be the fraction of scores expressing the degree of liking. This approach to analysis of data closely resembles that of the present study. Howgate concludes by highlighting the importance of the distribution of scores rather than relying on summary statistics such as medians. This practice provides more information.

Howgate also observes from a survey of papers relating to food science and technology, users of hedonic rating scales, almost without exception, assume an equal interval scale and process the data using parametric techniques which require the assumption that data have additive properties and that errors be normally distributed. The writer confirms his view as a result of literature surveys in relation to the present study.

Blakesley (1977) describes a Fortran computer program to perform an analysis based on nonparametric Friedman type rank statistics for a two-way classification. If an overall difference is established for multiple samples, a series of tests on sample pairs is carried out to determine which samples are statistically different. The author comments that a commonly employed statistical method of analysing sensory panel results for difference is a Fisher one-way analysis of variance or F test followed by the Duncan Multiple Range test. He does not consider this technique to be entirely adequate since it is unusual for/

for the same panellist to evaluate each sample thus violating the independence assumption for a one-way analysis of variance. Responses are usually on an ordinal scale. Even if it can be shown that they are normally distributed he claims that the usual F-test is known not to have the F-distribution. He recommends the use of the Friedman rank statistic which eliminates the problem of normal distribution. The Friedman statistic also allows association due to individual preferences between samples judged by an individual panellist.

As has been indicated, the Systems Analyst and Statisticians in the Data Processing Centre at QMC have always been concerned that judgements by an individual whether on the basis of hedonic scales, preference or difference between samples should be considered other than highly subjective. They are not convinced that tasters, particularly the less experienced, can be considered to be objective in their judgements. In making recommendations on experimental design and on the analysis of data they have always believed that the food scientist should use the techniques of the behavioural scientist in that the same or equivalent tests are used by both groups. Throughout these studies they urged a conservative approach. Most of the non-parametric tests which were used are only slightly less powerful than equivalent parametric techniques and are not open to assumptions regarding normality of distribution and independence of observations. At a practical level, observations cannot be considered to be independent when several samples are assessed together by the use of hedonic rating scales, in triangle tests and in paired comparisons. The Wilcoxon Matched-Pairs Signed-Ranks test was recommended in the comparison of correct identification of the odd sample in triangle tests on raw and cooked meats rather than the parametric t test. It may/

may be that as a result Type I errors may have occurred where results were marginal. Throughout, they urged this cautious approach to experimental programmes involving human responses and in doing so have acquired considerable background knowledge on sensory appraisal techniques.

It is accepted that those who use nonparametric techniques are at present in a minority. Many experimenters admit that they are more appropriate for analysing sensory appraisal data, particularly hedonic rating scales, but do not use them. However discussions have been initiated and statisticians are increasingly willing to co-operate with food scientists. So it may be that in the future existing computer programs for parametric statistics will be replaced by those which were used as the basis for the present study.

The author acknowledges the discussions, advice and co-operation of the QMC staff which has been received during the present study. Their assistance could not have been expected had their recommendations been disregarded. Analysis of data would otherwise largely have been restricted to the use of statistical tables rather than to the more wide ranging studies carried out. It should also be noted that discussions took place before the start of the trials and on their completion with the ARC statistician M M Franklin.

Perhaps it may be that those who have been using parametric statistical techniques in the analysis of data for so many years will eventually come to review their long established procedures. After selecting the sensory appraisal techniques to be used in the present study, as described in Chapter 3, methods used to analyse the data seem entirely appropriate in the College situation particularly as no statistical assumptions are violated. In any case if previous methodologies are uncritically followed/

followed progress in research may be slow as any inherent defects in experimental design will remain undetected.

5. Results of the Present Study

(i) Hedonic Rating Scales

Following the procedures indicated in Chapter 4, hedonic rating scales were used to determine the optimal internal temperature for roasted gigots and loins. These scales are widely used in other laboratories and are considered to have considerable advantages and attractions particularly for the less experienced. However, because most of the samples presented either hot or cold were so attractive to participants, they found it extremely difficult to assess their attitude to them. No doubt they would have been dismayed to learn how their performance varied in duplicated trials at the same session.

As a result, no clear cut preference emerged from the trials and extensive study of the data was required before deciding that, on balance, 75°C was the most preferred temperature for both gigots and loins. It was also concluded that there was no evidence that flavour differences exist between left and right joints from the same carcass and that acuity was not affected whether joints were tasted hot or cold.

The more structured hedonic rating scales used for assessment of tenderness, juiciness and flavour used by several groups of BSc students of Agriculture at ESCA were no more successful in identifying differences in samples cooked to internal temperatures of 70°C and 80°C except in relation to juiciness. These students usually perform well in other sensory appraisal tests.

In view of these findings two possibilities should be considered. In lamb as opposed to other carcass meats, differences between samples cooked to different internal temperatures may be so slight as to make it/

it difficult to indicate a different attitude to each sample even if more structured scales are used. This possibility requires further and very careful investigation possibly by the use of difference tests. Alternatively, it may be that hedonic rating scales cannot be used to demonstrate differences between samples unless they are very considerable. This was unlikely in the present studies. Extensive experience with other foods at QMC suggests that demonstrating statistically significant differences between samples by the use of hedonic rating scales is not likely to be as successful as when difference tests are used.

It would be of interest to determine what is implied by 'training' in laboratories where meats are assessed using hedonic rating scales. Pearson et al (1973) demonstrated comparable results from trained and inexperienced tasters. Park et al (1975) demonstrated good agreement between mean hedonic ratings of two inexperienced taste panels at North Ryde with the two Cannon Hill panels (experienced) who were more critical than the former and thus awarded lower scores. Preferences of the four panels were in agreement. Admittedly the quoted means give no indication of the distribution of responses. The two inexperienced panels may have shown just as great a variation in duplicated trials as in the Edinburgh studies. The power of hedonic rating scales should also be determined particularly in replicated tests and it is recommended that such studies should be undertaken as a matter of urgency in view of the present findings.

There is however no doubt that the continued use of this technique could not have been justified for the main trials which were to follow in the present study. It was considered prudent to devise a more precise experimental technique even if at a later stage further experiments

of/

of a different type would be required.

(ii) Comparing the Flavour of Grass and Rape Fed Lambs

In the first series of trials, no statistically significant differences in the aroma or flavour of the samples was demonstrated by the use of triangle (difference) tests. No consistent preference for samples from either feeding regime was demonstrated. Flavour profiles indicated that there were no flavour attributes common to samples from a particular feeding regime. Possible reasons for the failure to confirm the findings of Park et al (1972b) are indicated in context.

When the trials were repeated on lambs born in April/May 1978, there was some indication that there could have been slight flavour differences between roasted gigot and loin samples and that the flavour of the grass fed lambs was preferred. Aroma was not assessed in the second series of trials. In January 1982, ESCA students assessed the aroma of raw and cooked tissue of samples used in the second series of trials. Because they had been stored frozen since December 1978, it was not considered appropriate that the flavour assessments should be made. Despite prolonged storage, the defrosted samples showed no sign of rancidity but colour was not as good as in previous samples probably as a result of metmyoglobin formation. On this occasion, statistically significant differences in the aroma or raw and cooked tissue were demonstrated using triangle tests ($p<0.001$ and $p<0.05$ respectively). When the observed and expected frequencies were compared the same levels of significance were obtained. It thus seemed possible that differences existed between samples although the Kolmogorov-Smirnov test did not indicate that distribution of responses was other than binomial. The Wilcoxon Matched-Pairs Signed/

Signed-Ranks test indicated that subjects' performance did not differ when assessments were made on raw and cooked samples.

In reviewing these results, the prolonged period of frozen storage must be considered. Tissue from rape fed animals is likely to contain appreciable concentrations of glucosolinates. In poultry (Nute - personal communication) flavour defects in both eggs and muscle have been reported. Steedman et al (1979a and 1979b) and Yule and McBride (1979) indicate that rape fed poultry had less acceptable and sometimes more fishy flavour than controls. Nute considers (personal communication) that metabolism of glucosolinates may differ in avian species causing an accumulation of trimethylamine and other undesirable flavour compounds. Thus it may be that the differences detected by the ESCA students were a result of degradation of glucosolinates rather than of feeding regime per se.

In reporting on the feeding of rape in 1974, Wheeler et al record variation in the intensity of undesirable flavour which had been observed in studies subsequent to those reported in 1972. In experiments of 1969 and 1970 the rape cultivar used was Rangi. During one trial where the grazing period was only six days, rape fed lambs were much less acceptable and had an intense and unpleasant aroma and flavour.

This trial was subsequently repeated using the rape cultivar Aphid Resistant. The flavour of meat slaughtered directly off this rape was mild, to an extent that it was not possible to determine if withdrawal of the sheep to grass for seven or fourteen days pre-slaughter produced lamb of more acceptable flavour (Czochanza et al (1970).

Thus the possibility of certain rape cultivars inducing stronger flavour than others was investigated in a third trial. Taste panels found/

found no significant difference in intensity of flavour or in acceptability of meats grazed cvs. Rangi, Aphid Resistant or Giant Kangaroo. Rangi did not therefore produce the unacceptable aroma and flavour reported in previous trials.

Wheeler et al (1974) therefore concluded that whilst grazing sheep on rape can produce strong, characteristic flavour that this effect is not consistent. Cultivar may have influenced later results but it is suggested that nitrogen fertilizers may have induced the production of higher concentrations of glucosinolates and their metabolites in rape cv. Rangi in 1969 and 1970. This could have been the cause of flavour defects in lamb from this feeding regime.

Thus it would be reasonable to conclude as a result of the two Edinburgh trials, that if flavour differences do exist between grass and rape fed samples that they are so slight that they would be undetected by the average consumer.

(iii) Tasters Performance

In the trials to compare older turnip fed wethers with younger barley fed crossbreds (Chapter 7) a high proportion of tasters at both ESCA and QMC were able to detect aroma and flavour differences by the use of triangle tests. They were also consistent in their preferences. The majority of tasters preferred samples from the older lamb. Thus the tasters were considered to show aptitude for the testing procedures subsequently used in the grass versus rape trials and in the comparisons of the winter forage crops, the stubble turnips and grass silage and lucerne silage trials. The Kolmogorov-Smirnov test indicated that predominantly responses were not binomially distributed in the triangle tests on flavour. This confirms tasters' ability to detect flavour differences when samples such as those described in Ch. 7 are compared (in contrast to the other comparisons made in the present study. That in only/

only two of the preference trials did distribution of responses differ from binomial distribution is considered to have arisen as a result of the highly subjective nature of preference. It was thus concluded that the methodology devised was likely to meet the requirements of the present study.

(iv) Other Forage Crops

As explained in Chapter 10, the effects of feeding barley, swede, cabbage, stubble turnips and lucerne were studied as part of teaching programmes at QMC and ESCA. Comparisons between two feeding regimes, indicated in Chapter 9 were:

Barley versus Swede

Barley " Cabbage

Cabbage " Swede

Stubble turnips versus Stubble turnips (left and right sides)

Grass versus Grass silage

Grass silage versus Lucerne silage

There was little difference in preference for either sample in the first four trials which were carried out at QMC. The flavour of the swede fed samples was preferred to the barley fed samples in two of the four trials but in only one of these trials was a statistically significant difference in flavour established by triangle tests. In three of the four trials, detectable flavour differences were demonstrated between barley and cabbage fed lambs, although in only one trial did this difference affect preference. No detectable difference in flavour was detected in samples from cabbage and swede fed lamb, nor was there preference for lamb from either feeding regime. The anomalous results in the Stubble Turnip trials are considered in context. Differences in aroma of raw and cooked samples/

samples were few and showed no consistent pattern except for the comparisons between the cooked samples from cabbage and swede fed lambs. Results of Kolmogorov-Smirnov tests confirmed that differences in flavour were present between barley and cabbage fed samples. In both the barley versus cabbage and cabbage versus swede trails, cabbage fed samples were preferred. Barley fed lambs were less frequently preferred than either meats from swede or cabbage fed lambs. The Wilcoxon Matched-Pairs Signed-Ranks tests indicated that there was no difference in subjects' ability to detect differences in aroma when samples were assessed raw and cooked.

In 1980, ESCA students compared samples from grass and grass silage fed lambs. Flavour preference for grass fed lambs was demonstrated ($p<0.005$). This preference was not maintained when based on general acceptability which is a much less precise characteristic which takes into account such factors as appearance, texture including tenderness, juiciness and mouthfeel, in addition to flavour. The purpose of these trials was to determine the effects of feeding silage as opposed to grass in preparation for the grass silage and lucerne silage trials which were to follow.

These trials took place in January 1981 at ESCA. There was no detectable difference in the aroma of the cooked lean or lean plus fat samples of either feeding regime. No statistically significant differences in flavour were established. This was not a matter of surprise since aroma is a major component of flavour. There was however a tendency for the flavour of the grass silage fed samples to be preferred in paired comparison (preference) tests where the basis of preference was specified as flavour.

Thus as a result of trials described in Chapters 6, 8 and 9, careful/

careful experimentation has demonstrated that the fattening regimes for store lambs used in the 1979 ESCA trials have exerted little effect on lamb aroma or flavour. Whilst it could be considered desirable to repeat these experiments, the present results suggest that it is most unlikely that their findings would differ. From these trials and those described in Chapter 5, it is suggested that lamb is a very variable meat which Edinburgh tasters find very difficult to assess. In most trials test and control samples were almost indistinguishable with lamb flavour overriding any flavour differences which could have arisen as a result of feeding regime. It seems very unlikely that the consumer would reject samples from any of these store lambs.

In relation to the lucerne silage trials, it should be noted that Park et al (1972a) reported that significant ($p<0.01$) interaction was found between pasture type and condition of pasture. Lamb was less acceptable when lucerne plants were leafy and growing vigorously than when lucerne growth was poor and the plants stemmy. Thus the effects of feeding lucerne cannot be considered to be consistent particularly if lambs are transferred to grass for a week before slaughter. This is in agreement with the findings of Shorland et al (1970) and Czochanza et al (1970). The conversion of lucerne to lucerne silage in any case introduces further variables.

(v) Protected Lipid Supplements

The change in method of recording flavour acceptability in these trials to very poor through to very good has already been noted. As indicated in Chapter 1, pasture fed lambs were preferred to those fed protected lipid supplements. Alterations occur as a result of length of feeding and nature of supplement since sunflower seed was criticised less than safflower seed. Wright et al (1974) found that whilst flavour/

flavour was different it was not unacceptable. Garrett et al (1976) confirmed that 18 : 3 fatty acid concentration in fatty tissue increased but palatability studies were not carried out. Adverse comments have been made mainly by Australian workers. They consider (Ford and Park 1980) that the oily flavour detected may be preferred by North Americans and Japanese accustomed as they are to the oiliness of pork and poultry. Such flavour could therefore be advantageous.

(vi) Objective Data

This data was collected as a contribution to fundamental knowledge and to determine if any of these objective parameters could be linked to predictions of flavour.

The following conclusions were reached:

1. Although there are considerable differences in initial weights of left and right gigots of standard anatomical location these are not of statistical significance. Although these initial weight differences are reflected in total and evaporative weight losses, differences between right and left sides are not statistically significant. In future experiments it may not therefore be necessary to compare joints from the same side of carcasses provided that anatomical location is standardised.

2. When comparing total percentage weight losses (similar to those of Oliver et al (1970) and Kemp et al 1972) irrespective of feeding regime, there are statistically significant differences in weight losses between gigots and loins. Total weight losses are lower for loins. Whilst it is clear that in most laboratories meats of standard anatomical location are used in comparisons, this finding confirms the importance of following this practice.

3. There are also statistically significant differences between evaporative/

evaporative losses between gigots and loins. Loins show greater uniformity of evaporative weight loss than gigots. This finding again emphasises the necessity of using standard samples for experiments. Both in relation to total and evaporative losses, variations indicate that as large a group of samples as possible should be used since such intrinsic variation exists between samples.

4. There are statistically significant differences in initial pH values of gigots and loins. The loins have much lower pH values. Loins showed less variation in initial pH than gigots. This difference could affect changes induced by the cooking process particularly those relating to tenderness and juiciness.

5. There are statistically significant differences in cooked pH values of gigots and loins. Again, loins showed lower pH values and less variation than gigots.

6. Despite statistically significant differences in initial and cooked pH values, changes in pH in response to the cooking process showed no statistical differences ($p<0.30$). Changes in pH are only slightly more uniform in loins than gigots.

6. Left and right gigots were compared in relation to response to the cooking process. There was no statistically significant response between joints from the left and right sides of the same carcass ($p<0.30$ and $p<0.20$ (1df) for gigots and loins respectively). Changes were however more uniform in left side than right side joints.

7. Loins are concluded to be less inherently variable than gigots and may therefore be better suited for experimental studies. It is nevertheless important to take into account observed differences in total and evaporative weight losses as a result of the cooking process and/

and statistically significant differences in initial and cooked pH values when comparisons are made.

Although considerable importance is attached to comparisons between joints of standard anatomical location, to the extent that only a single muscle may be used in comparisons, it is concluded that in future studies greater emphasis should be placed in studying inherent differences between cuts of meat or muscles selected for testing. These findings present quantitative evidence of the need to standardise location of meat samples as well as the need to investigate the nature of possible differences between them. This practice has been followed in many laboratories but from literature studies it is impossible to judge whether this approach has been intuitive or based on quantitative assessments.

(vii) Comparing the Aroma and Cooked Samples and a Study of the Contribution of Fat to Flavour

In Chapter 7, tests to monitor subjects' performance are described; statistically significant differences in the aroma of raw and cooked gigot samples were demonstrated ($p<0.00$ and $p<0.05$ respectively). For loins, although there were no detectable differences in the raw samples, statistically significant differences in cooked samples were shown ($p<0.01$). Results from loins may have been affected by the small panel size since there were only six participants. When the results for all gigot and both loin trials are combined, the possibility of achieving the difference by chance alone is zero (Table 7.9). In four of the seven triangle tests for flavour results are not binomially distributed. The same applies to two of the six trials relating to triangle tests on the aroma of raw samples and to two of the seven trials relating to paired comparisons/

comparisons (Table 9.10). It is therefore clear that subjects found far more readily detectable differences between these two samples than in all the other trials of the present study. The Wilcoxon Matched-Pairs Signed-Ranks test indicates that subjects do not detect differences any more readily in raw than in cooked samples (Table 7.11).

This is a finding of considerable importance. If differences in aroma are detected just as readily in raw samples, it is likely that the same result would have been obtained by testing small samples of the raw tissue from the appropriate anatomical locations rather than as in these experiments, cooking the samples. This would not of course apply if flavour preferences or flavour differences were to be determined. In future experimental programmes, it is suggested that before major trials involving cooked samples are undertaken, the feasibility of sampling the raw carcass initially should be investigated. If this sampling procedure were to prove successful, there could be considerable financial and economic implications.

The results discussed above should be compared with those set out in Table 11.1. In all the trials relating to the fattening regimes for store lambs in 1977, with only two exceptions subjects' performance in triangle tests was the same in raw as in cooked samples. These two probably anomalous results are discussed in context in Chapter 9. This also applies to trials in which statistically significant differences in aroma were detected. This confirms the findings of Chapter 7 and indicates that further investigations would be of great value.

The contribution of fatty tissue to lamb aroma and flavour was studied in experiments carried out by students at ESCA. When samples of/

of grass silage and lucerne silage were assessed for aroma, there was no statistically significant difference established. The presence of fatty tissue did not influence this result (Table 11.2). Similarly species difference in the aroma of cooked lamb and beef were demonstrated in the presence or absence of fatty tissue. These differences were of statistical significance. Although results must be regarded with caution in view of possible distracting effects of texture, the flavour of samples of lean tissue of lamb and beef was readily distinguished in triangle tests ($p<0.001$). For lamb it would thus appear that sufficient of the water soluble phospholipid fraction is present in the perimysium to confer characteristic species aroma and flavour. during the cooking process (Rhodes 1973; Rhodes 1978; Mottram and Edwards 1983).

This finding does not correspond to some of the tests which were carried out with pork. When the aroma of cooked fatty tissue from young and mature pigs was compared using triangle tests, detectable differences were present in both raw and cooked samples ($p<0.01$). As with lamb, the cooking process did not affect this result. There were also detectable flavour differences in the flavour of lean tissue ($p<0.01$) as assessed by triangle tests. In subsequent tests, using younger pig and boar meats, the presence of fatty tissue increased subjects' ability to detect differences between the aroma of the two samples. The value of χ^2 derived from comparing correct and incorrect responses was 23.06 indicating that fatty tissue made a very highly significant contribution to subjects' performance. This is in contrast to the species identification trials of lamb and beef where subjects' performance improved if only lean tissue was present. This result was not however of statistical significance (χ^2 value was 3.57 (1df)/

(1df) $p<0.10$). Further results of this series of trials are given in Chapter 11 but are not of relevance in comparing the contribution of fatty tissue in cooked lamb with that of cooked pork samples. It is clear that lean cooked tissue of lamb possesses aroma and flavour characteristic of species provided that fatty tissue is not removed before cooking. These results suggest that further experimentation to determine the effects of fatty tissue on lamb flavour would be of interest.

6. Conclusions

Results of trials described in this study indicate that tasters in Edinburgh perform better in experiments designed to detect difference and establish preference as described in Chapters 6, 7 8 and 9 than they do in hedonic rating scales. This finding was unexpected in view of techniques so widely used in other laboratories. It could however be argued that scores derived from hedonic rating scales are not necessarily directly related to preference for or difference between samples although they are as frequently used for this purpose. Further comparisons of testing procedures are therefore a priority.

Although difference and preference testing are usually considered to be a preliminary to tests of a more descriptive type, this study of the effects of these fattening regimes on the flavour of samples from store lambs does not indicate that such additional tests would be of value. Consumers are unlikely to reject samples of store lambs from any of the fattening regimes studied. In noting the variation in gigot and loin samples and the statistically significant differences between them, the value of further studies

of/

of the nature of these differences is indicated. The fact that differences in aroma of lamb are detected just as readily in raw as in cooked samples has considerable economic and practical implications and it is suggested that this important finding should receive further study. It would also be of interest to make a further study of the contribution of fatty tissue lamb flavour particularly as sheep meat seems to show such intrinsic variations and lamb odour and flavour are sufficiently intense to mask any slight differences in flavour resulting from feeding regime. Throughout the study it was demonstrated that roasted lamb samples have quite different sensory attributes to those of beef and pork. It would be of value to investigate the reasons for these differences more fully.

The experimental results described in this study have demonstrated the need for further investigations particularly in relation to the causes of variation in sheep meat samples. The experimental techniques which have been devised have been shown to be appropriate in a College situation where tasters can participate only in a limited number of trials and where the sensory appraisal of meats is only a small part of experimental programmes. It is hoped that it will be possible to continue with andextend such collaborative experimental programmes in the future.

ACKNOWLEDGEMENTS

The author acknowledges with gratitude the assistance, encouragement and support of:

Professor J H D Prescott, Internal Supervisor.

Mrs M Doran, Chief Technician, QMC.

Staff of ESCA and QMC.

Mr J Fraser, Carcass Evaluation Unit, ESCA.

Mr J FitzSimons, Farms Division Office, ESCA.

Mr M M Franklin, ARC Statistician.

Mr I Garland, Lecturer, QMC.

Miss E Graham, Graphics Artist, QMC.

The Librarians, QMC.

Mrs G Mackie, Senior Technician, QMC.

Mr A McGregor, Printer.

Miss A Milne, Technician, QMC.

Miss E M Morrison, Director, Information Technology, formerly Data Processing Centre, QMC.

Mrs L Mungall, formerly Senior Technician, QMC.

Tasters from ESCA and QMC.

Mrs M P Woods, Lecturer, QMC.

R, E C and M.

and Mrs I Rooney who typed this thesis.

APPENDIX

TABLE 1.1

A Comparison of Daily Average Nutrient Intake in the United Kingdom and Scotland

Nutrient	United Kingdom	Scotland
Energy Kcal	2,250	2,250
MJ	9.5	9.5
Protein g	73.4	73.6
Fat g	106	103
Carbohydrate* g	268	275
Thiamin mg	1.22	1.19
Riboflavin mg	1.90	1.80
Nicotinic acid equivalents mg	30.6	30.1
Iron mg	11.0	11.3
% energy derived from:		
Protein	13.0	13.1
Fat	42.4	41.2
Carbohydrate*	44.6	45.8

*as monosaccharide.

TABLE 1.2 Contribution Made by Carcass Meats to Daily Energy, Protein, Fat and Iron Intakes

	Energy			Protein		Fat		Iron		Saturated Fats		PUFA	
	Kcal	MJ	%	g	%	g	%	mg	%	g	%	g	%
Beef and Veal	69	0.29	3.0	5.9	8.1	5.1	4.8	0.7	6.0	2.1	4.4	0.2	1.9
Mutton and Lamb	43	0.18	1.9	2.2	3.0	3.8	3.6	0.2	1.6	1.8	3.9	0.2	1.7
Pork	38	0.16	1.7	2.1	2.8	3.4	3.2	0.1	0.9	1.3	2.8	0.3	2.4
Total	149	0.63	6.6	11.2	13.9	12.3	11.6	1.0	8.5	5.2	11.1	0.7	6.0

where % refers to per cent of total daily intake and PUFA = polyunsaturated fatty acids

Source: National Food Survey 1979

TABLE 3.1

QUEEN MARGARET COLLEGE

MEAT RESEARCH PROJECT - TASTING TESTS - ROAST LAMB

To: ALL TESTERS - INSTRUCTIONS

Please will you taste all samples of roasted lamb from each batch. Decide which of the following descriptions is most appropriate to the FLAVOUR of a particular sample:

Like extremely	(1)
Like moderately	(2)
Like slightly	(3)
Neither like/nor dislike	(4)
Dislike slightly	(5)
Dislike moderately	(6)
Dislike extremely	(7)

The description which most closely matches your response to the flavour of the meat sample should be scored as indicated in brackets:

e.g. description "Like moderately", therefore score (2)

description "Dislike slightly", therefore score (5)

Retasting is allowed but too much retasting may be confusing. Use water as a palate cleanser between samples. Record your findings on the sheets provided.

This type of assessment is known as HEDONIC RATING SCALE.

Date .....

Times .....

Place .....

Thank you for your co-operation. Please continue to participate in these testing procedures.

TABLE 3.2

QUEEN MARGARET COLLEGE - SENSORY APPRAISAL OF MEAT SAMPLE

NAME ..... DATE .....

TEST TYPE ..... TEST NUMBER .....

TASTER NUMBER

SAMPLE				
SCORE ON HEDONIC SCALE				

TABLE 3.3 (Available to tasters during all sessions)

QMC/ESCA RESEARCH PROJECT

HEDONIC SCALE RATINGS - INSTRUCTION SHEET

Taste each of the samples provided. Decide how much you like or dislike the flavour of each sample. Use the scale below to indicate your opinion of the flavour of the samples.

	Score
Like very much	1
Like moderately	2
Like slightly	3
Neither like nor dislike	4
Dislike slightly	5
Dislike moderately	6
Dislike very much	7

Retasting is allowed but too much retasting may be confusing. Palate cleaners should be used.

PLATE 5

Alternative Presentations of Samples in Paired Comparison (Preference) tests. (as described in Chapter 4)

PLATE 6

Alternative Presentations of Samples in Triangle Tests. (as described in Chapter 4)


201
201
855
855
725
725

TABLE 4.1

P A P E R C

ANALYSIS OF RESULTS - ROASTED LAMB JOINTS

DOREEN A PARRY QUEEN MARGARET COLLEGE, EDINBURGH 4

OCTOBER 1977

Carcass:
Date of Cooking:
Date of tasting:
Type of Joint:
Sample Tasted: Side of Carcass:

Oven Temperature: 177°C(350°F) Internal temperature attained: Variable

RESULTS	SAMPLES		Program: File:	DPA
Recording Internal temp. attained °C	70	75	80	85
Initial pH meat				
pH cooked meat				
Weight raw meat g.				
Weight roasting tin g.				
Weight cooked meat g.				
Total weight loss g.				
% Weight loss				
Wt. roasting tin + fat g.				
Fat loss g.				
% Fat loss				
Water loss g.				
% Water loss				
Cooking period (mins)				

TABLE 4.2

QUEEN MARGARET COLLEGE

FOOD SCIENCE

Oven temperature:

Internal temperature attained:

Data / Sample & Cut														
Initial pH meat														
pH cooked meat														
Weight raw meat g.														
Weight roasting tin g.														
Weight cooked meat g.														
Total weight loss g.														
% weight loss														
Wt. roasting tin + fat g.														
Fat loss g.														
% Fat loss														
Water loss g.														
% Water loss														
Cooking period (mins)														
Flavour preference														

TABLE 4.3

QUEEN MARGARET COLLEGE

MEAT RESEARCH PROJECT

(In conjunction with BSCA)

Preliminary Testing Procedures

Name of tester:

Date of test: 18/3/77

After completing these tests, please leave this form in the tray provided.

1. Primary Tastes

There are four primary tastes - sweet, sour, salty and bitter. Each is identified by taste buds located at specific areas of the tongue.

Four of the flasks contain solutions to enable recognition of one of the primary tastes. The fifth contains tap water.

Taste the contents of each of the flasks. Identify the taste of the four solutions.

Flask	Primary Taste
1	Salty
2	Sour
3	Sweet
4	water
5	bitter.

Retasting is allowed but do not retaste too much as this may be confusing. You will find it helpful to rinse your mouth with water in between tastings.

2. Mixtures of Primary Tastes

You are provided with three solutions. Each solution is a mixture of 2 or 3 primary tastes. (sweet, sour, salty and bitter).

Taste each solution. Identify the primary tastes present in each. Make

a tick in the appropriate space in the table below when you recognise the presence of one of the primary tastes.

Solution	Sweet	Sour	Salty	BITTER
A	✓	✓		✓
B			✓	✓
C	✓	✓	✓	

Great care is required in this test. It may be necessary to taste each solution more than once to identify either the two or three primary tastes.

3. Ranking Test

Three sugar solutions, L, B and V are provided. Taste each of them and rank in order of increasing concentration.

	Solution
Lowest Concentration	B
Medium Concentration	L
Highest Concentration	V

Retasting is allowed but you are advised not to retaste too much as this may be confusing.

4. Odour Recognition

Contained in the bottles are 20 fairly common odorants - types of smells with which you should be familiar.

Please describe the smell of each if you know the chemical name, please state it, otherwise describe the smell. Even if you cannot remember what the smell is, if you can remember where you encountered the particular smell please say so, eg.

Hospital, dentist, florist, grocer, chemist etc.

There is something in every bottle - it is better to identify those with which you are familiar first and then go back to the ones which you did not identify immediately.

If you think a bottle has no smell, pause for a short time as you may be suffering temporarily from odour fatigue.

Please write <u>something</u> for each odorant.

Bottle No.	Odour/association. Please describe	
1	Cod Liver Oil	3
2	menthol	5
3	pear drop odour	4
4	gunk	5
5	vanilla	5
6	tobacco.	2
7	camphor - reminscent of having a cold	3
8	gums powder	5
9	fishy smell that catches back of nose	2
10	nutmeg	5
11	ammonia - faint smell of cleaning materials	
12	almond essence - marzipan smell	5
13	Lozenges ?	cough mixture smell 4
14	Cloves	5
15	Mothballs	4
16	Lemon essence - not quite a pear drop smell	5
17	Cigarette Ash - Stale smoke smell	5
18	Oil paints ? - Some type of remover	3
19	dettol - antiseptic smell	3
20	Coconut. smell very faint, nutty	5

83%

Thank you for taking part in these tests. As well as tasting meats, there will be additional tests of this kind from time to time.

I look forward to meeting you and to discussing our findings.

Mrs D A Parry
Senior Lecturer in Food Science
and Nutrition

KEY TO TESTS

1.	Solution	1	0.2% sodium chloride w/v
		2	0.06% citric acid w/v
		3	2.0% sucrose w/v
		4	0.07% caffeine w/v
		5	Deionised water
3.	Solution	B	8% sucrose w/v
		L	10% " "
		V	12% " "
4.	Odorant	1	Linseed oil
		2	Menthol
		3	Iso-amyl acetate
		4	Garlic
		5	Vanilla
		6	Sage
		7	Vick
		8	Curry power
		9	Acetic acid
		10	Nutmeg
		11	Ammonia
		12	Benzaldehyde
		13	Oil of aniseed
		14	Eugenol
		15	Naphthalene
		16	Citral
		17	Tobacco
		18	Phenol
		19	Pine disinfectant
		20	Oil of coconut

Results of Test 2 were omitted.

TABLE 4.4

QUEEN MARGARET COLLEGE

RESEARCH PROJECT - MEATS

(In combination with ESCA)

Dear

Thank you very much indeed for volunteering to take part in this Research Project. Perhaps when you hear more about what is involved you will encourage others to participate! The more participants there are, the more your personal commitment will be reduced!

Results from large numbers of tasters are of greater accuracy. It is also of great importance that our ability to carry out such research projects at Queen Margaret College should be recognised.

Most of our work will be concerned with tasting roasted lamb samples (hot and cold) or 'sniffing' heated cooked samples to detect possible differences in aroma. I would hope that some - if not all of you! - could become trained taste panel members. Occasionally, we may be asked to taste/'sniff' other meats.

In addition, I hope that you will agree to take part in experiments which are designed to select - with great stringency - traditional taste panel members. As a result of research projects already carried out in College for meat tasting/flavour assessments these criteria do not always apply here! Some of my best performers would have been rejected at the first stage of selection as panel members!

I look forward to contacting you again in the near future.

You will be reimbursed for any "out of pocket" expenses in attending tasting sessions and it is hoped that from time to time tasters will meet together to discuss progress in a more social setting.

With many thanks and good wishes.

Yours sincerely

MRS D A PARRY
Senior Lecturer in Food Science and Nutrition

TABLE 4.5

QUEEN MARGARET COLLEGE

RESEARCH PROJECT : MEATS

(In combination with ESCA)

Preliminary Testing of Tasters

Four tests are normally carried out to select traditional taste panel members. The Meat samples are not yet available, but I should be most grateful if you would carry out these four tests during the week beginning March 13th at one of the times indicated below. They should be carried out in the same order as they appear on the recording sheet provided. Allow yourself approximately 20-30 minutes.

If you are a smoker, do not smoke for at least half an hour beforehand. Hands should first be washed in the unperfumed detergent provided in dispensers in College cloakrooms. Perfumed cosmetics are best avoided that day. Most people find that taking a sip of water between tests helps to avoid confusing the palate. Finally, do not allow yourself to be influenced by the reactions of other participants!

ROOM 429 (Science Department)

on	TUESDAY	14th March	9.15am - 1.15 pm.
	WEDNESDAY	15th March	" "
	THURSDAY	16th March	" "
	FRIDAY	17th March	All day

Thank you for your co-operation.

MRS D A PARRY
Senior Lecturer in Food Science and Nutrition.

TABLE 4.6

QUEEN MARGARET COLLEGE

TASTERS WANTED!

Staff and Students of Queen Margaret College have often been involved in assessing the eating quality of foods. Volunteers are required to taste samples of roast lamb on Wednesdays or Thursdays of weeks 3, 4 and 5 of this term (12.15 – 1.30 p.m.). No previous experience is required but it is most important that you attend all three tasting sessions.

Please sign the list in the pantry if you would like to take part in this important series of tests.

Mrs D A Parry
Department of Science and Dietetics

TABLE 4.7

QMC/ESCA RESEARCH PROJECT

MEAT TASTING

Dear

I hope that you will be able to participate in the tasting session, details of which are given below.

Thank you for your co-operation.

Yours sincerely

MRS D A PARRY
Senior Lecturer in Food Science and Nutrition

Date .....

Times .....

Location .....

TABLE 4.8

QMC/ESCA RESEARCH PROJECTS

Report on Tests of 1st March 1978

1. Number of Tests Carried Out

Just right	7 participants
Too many	2 "
No response	6 "

Thus to limit attendance at this series of tests the experimental programme will remain much the same.

2. 'Guessing' Your Response

Always give a reply however difficult this may be. Scores awarded by chance alone are allowed for in analysing the data.

3. Standardising Tasting Procedures

Try to chew samples thoroughly and roll them in the mouth to achieve maximum flavour perception. Differences or preferences are more readily identified if tasting techniques are standardised.

4. Triangle Tests - Aroma

Always carry out the tests on the tubes in each beaker at the appropriate assessment section set aside and not beside the heater.

5. Swallowing Samples

After tasting a piece of meat it can be discarded by the use of paper towels provided if you wish.

6. Completion of this Series of Tests

There will be two more sessions. The first will be on March 15th and the second on March 22nd.

Once again, thank you for your help.

DOREEN A PARRY
Queen Margaret College

8.3.78

TABLE 4.9

QUEEN MARGARET COLLEGE

QMC/ESCA RESEARCH PROJECTS

MEAT TASTERS

RESULTS OF RECENT SERIES OF TESTS

The roasted joints we compared were so very much alike that:

1. In paired comparison (preference) tests, no preference for either batch of samples was shown.

2. Identification of the odd sample in triangle tests on the basis of flavour was difficult if not impossible! This again suggests that samples were almost identical.

3. The same situation arose with the triangle (aroma) tests.

It is essential these results - which are encouraging to lamb producers - should be checked by a second series of tests. Three sessions have been arranged as indicated below.

Your participation in the previous sessions has been greatly appreciated. We hope that you will be able to take part in this new series. Look out for the yellow notices to remind you!

DOREEN A PARRY

To:

Dates	Times	Place
Wednesday 31st May	12.30 - 1.30	Botany South
Wednesday 7th June	"	"
Wednesday 14th June	"	"

TABLE 5.1 Response Patterns - Frequency Distributions 1 - 7

				A 70°C								B 75°C								C 80°C								D 85°C							
File	Cut	T°	N	Median	1	2	3	4	5	6	7	Median	1	2	3	4	5	6	7	Median	1	2	3	4	5	6	7	Median	1	2	3	4	5	6	7
B	G	C	28	3	6	6	4	3	3	4	2	3	4	8	6	4	6	0	0	3	1	6	8	5	6	1	1	3	4	9	5	3	4	2	1
C	"	C	"	3	6	5	10	3	3	1	0	2	6	10	5	5	0	2	0	2	0	17	4	4	2	1	0	3	4	4	10	6	3	1	0
E	"	H	11	5	0	1	2	2	5	1	0	3	0	1	7	1	2	0	0	4	0	2	3	2	2	2	0	4	0	1	2	6	1	1	0
G	"	"	"	5	0	1	3	1	3	2	1	2	0	6	0	2	1	2	0	4	1	0	4	1	3	2	0	3	0	4	2	1	1	2	1
F	"	H	12	4	1	1	3	2	3	1	1	3	3	1	3	3	2	0	0	3.5	1	2	3	2	3	0	1	3	2	2	3	3	0	2	0
H	"	"	"	3.5	2	2	2	2	4	0	0	3	1	4	3	1	1	2	0	4.5	0	1	3	2	3	2	1	3	1	1	7	0	3	0	0
I	"	C	12	4	1	4	1	0	2	3	1	4.5	0	3	2	1	3	3	0	4	2	1	2	2	2	1	2	4.5	0	2	3	1	1	2	3
J	"	"	"	3	1	3	4	1	0	1	2	4	2	2	1	3	2	2	0	3.5	0	4	2	4	2	0	0	4	1	3	1	4	2	1	0
A	L	C	29	4	2	2	4	8	9	2	2	3	5	5	7	4	5	2	1	3	2	6	7	8	1	3	2	4	1	6	7	6	3	3	3
D	"	C	27	4	2	1	7	7	6	2	2	3	2	8	8	4	5	0	0	3	2	8	7	5	4	1	0	3	3	7	8	6	1	1	1
K	"	H	16	3	1	4	7	3	0	0	1	3	0	4	5	4	2	1	0	4	1	4	2	3	4	1	1	3.5	2	2	4	3	5	0	0
L	"	"	"	2	6	3	3	2	2	0	0	2	3	7	3	1	1	1	0	3	4	3	3	2	3	1	0	2.5	6	2	2	3	2	1	0
M	"	C	15	4	0	5	2	3	3	2	0	4	2	3	1	5	3	1	0	4	1	4	2	4	3	1	0	3	0	5	3	4	2	0	1
N	"	C	16	3.5	0	3	6	1	3	3	0	3.5	1	0	7	2	4	1	1	4	0	2	4	5	3	2	0	4	2	1	4	4	3	2	0

Table 5.2 Spearman Rank Correlation Coefficient - Results

(Identified Individual Scores)

1. Consistency of Judgements

		Files LA*	n	>0.05	1.0	-	->0.5	%>0.5	%-	%->0.5
E-H (L)	C	MN	15	10	1	2	0	67	13	0
	H	KL	16	4	2(-)*	6	3	25	38	19
A-D (G)	C	BG	26	8	1	8	5	31	31	19
	H	EG	11	3	1	2	0	27	18	0
E-H (G)	C	IJ	12	3	1(-)*	6	3	25	50	25
	H	FH	12	3	1	4	1	25	33	9
			92	31				34		

2. LHS v RHS and C v H

(L) A-D	AD	28	9	0	9	3	32	32	11
(L) E-H	MK	15	6	0	3	3	40	20	20
"	ML	15	3	0	9	5	20	60	33
"	NK	16	3	1	7	4	19	44	25
"	NL	16	3	1(-)*	9	5	19	56	31

(G)/

Table 5.2 (contd)

	Files LA*	n	>0.05	1.0	-	->0.5	%>0.5	%-	%->0.5
(G) A-D	BE	8	1	0	4	3	13	50	38
"	BG	8	2	0	2	1	25	25	13
"	CE	10	3	0	3	2	30	30	20
"	CG	10	3	0	5	5	30	50	50
(G) E-H	IF	12	3	1	6	5	25	50	42
"	JF	12	6	0	2	1	50	17	8
"	IH	12	6	1	5	2	50	42	17
"	JH	12	4	1(-)*	7	5	33	58	42
		84	28				33		

TABLE 5.3 Duplicates of Some Samples

Spearman's Rho Values - Table 6.18

Taster No.		LC MN	LH KL	GC BC	GH EG	GC IJ	GH FH	AD	LHS RHS MK	ML	C H NK	Comparisons NL	BE	BG	CE	CG	IF	JF	IH	JH
7	1			-				-												
8	2			0.1				0.8												
14	3			-				-												
15	4			0.5				0.6												
36	5			-				0												
37	6			0.8				0.2												
41	7	0.6	0.9	0.3				0.5	0.4	0.5	0.9	0.9								
45	8		0.8	1.0*				-			-	-								
90	9	0.5	-					0	0.7	-	0	0								
93	10			0.3				0												
94	11			0				-												
95	12	-	-	0.6	1.0*	0	0.4	-	0.6	0.3	-	0.8	0	0	0.6	0.6	-	0.8	0.5	0.8
96	13	0.3	-					0.5	-	0.8	-	0.8								
97	14	0.3	0.8		0.3	-	1.0*		-	-	-	-			-	-	-	0.8	-	0.8
98	15	0.8	-						0.8	-	1.0*	1.0								
99	16	0.9	-		0.6	-	0.2	0.2	0	-	-	-			0	0.8	-	0.2	0.9	-
100	17	0.9	-	0.2	0.3	0.9	-	0.5	0.2	-	0.4	0.1	-	-	-	-	1.0*	0.9	-	-
102	18			-	0	-	0.5	0					-	0	0.9	0	0.5	-	0.8	0
103	19	0.6	0.2	0.7	-	0.9	0.2	0.4	0.2	0.2	-	-	0.3	0.7	-	0.8	0.9	0.7	-	-
104/																				

- = negative correlation

TABLE 5.3 (Contd)

Taster No.		MN	KL	BC	EG	IJ	FH	AD	MK	NK	NL	BE	BG	CE	CG	IF	JF	IH	IH	JH
104	20	0.5	0						0.6	-	0.9	0.2								
109	21			-				0.7												
110	22			0.2				-												
111	23			0.2				0.9												
112	24	0.8	0.3	0.9	0	0.8	0.3	-	0.8	-	0.3	-	0.2	-	0	-	0.3	0.8	1.0*	0.8
113	25	1.0*	0.7						0	-	0	-								
114	26			0.8	0	-	-	0					0.5	0	0.3	-		-	0.5	0.5
116	27	-	0.3						0.7	0.2	0	-					-			
117	28			0.2				-												
118	29	0	0	-				0	0.3	-	0	0								
119	30			0.9		0.2	0	0									0	0	0.3	-
122	31			-	-	-	0.5	0.3					-	0.5	0.4	-	-	0.5	0.5	-
123	32			0.1	0.7	-	-	0.9					-	0.2	0.6	0.9	0.2	0.3	-	-
126	33	0.6	-			0.3	-		-	0.5	-	-					0.8	0.3	-	-
127	34			0.1	0			0.5												

Table 5.4

```
READY
999 FILES LAA
RUNNH
INPUT N,K
 ? 29,4
```

L (C) A - D (LHS)

THE DATA TABLE TO BE ANALYSED IS AS FOLLOWS.

EXPT 1 29 TASTERS	70 C	75 D	80 A	85 B
7 *	4	3	2	2
8 *	6	5	6	6
14 *	4	6	3	5
15 *	2	1	3	5
36 *	5	2	2	3
37 *	6	5	4	2
41 *	5	3	2	3
45 *	4	3	4	4
90 *	3	2	1	6
93 *	7	3	7	3
94 *	4	6	4	3
95 *	2	2	3	3
96 *	5	7	7	7
97 *	3	4	6	7
99 *	4	3	3	4
100 *	4	3	3	2
102 *	5	5	3	3
103 *	1	4	5	6
109 *	5	4	4	2
110 *	5	2	2	4
111 *	4	2	2	4
112 *	5	3	4	2
114 *	7	5	4	4
117 *	3	5	6	2
118 *	5	4	4	5
119 *	1	1	1	1
122 *	4	1	2	7
123 *	5	1	3	4
127 *	3	1	4	3
MEAN	4.2	3.3	3.6	3.9
STAN. DEV.	1.5	1.7	1.6	1.7

```
COL.  1 TOTAL =  87.5
COL.  2 TOTAL =  61
COL.  3 TOTAL =  67
COL.  4 TOTAL =  74.5

CHI R SQUARE =  8.1
```

* (> 7.85 for 3 df)

Table 5.5

LAMB TRIALS - JAN 78

TASTER NUMBER	SAMPLE A	B	C	D
7 *	4	3	2	2
8 *	5	5	6	6
14 *	4	6	3	5
[illegible] *	2	1	3	5
36 *	5	2	2	3
37 *	6	5	4	2
41 *	5	3	2	[illegible]
45 *	4	3	4	4
93 *	3	2	1	[illegible]
97 *	7	[illegible]	7	[illegible]
[illegible]4	4	6	4	3
95 *	2	2	3	5
96 *	5	7	7	7
97 *	3	4	[illegible]	[illegible]
99 *	[illegible]	[illegible]	[illegible]	4
100 *	[illegible]	3	[illegible]	[illegible]
101 *	5	5	3	3
[illegible] *	[illegible]	4	5	6
102 *	6	4	4	2
110 *	5	[illegible]	[illegible]	4
111 *	4	2	2	[illegible]
112 *	3	[illegible]	[illegible]	3
114 *	[illegible]	[illegible]	4	4
117 *	[illegible]	[illegible]	6	2
118 *	6	4	4	[illegible]
119 *	1	1	1	1
12[illegible] *	[illegible]	1	2	7
1[illegible] *	6	1	3	4
127 *	[illegible]	1	4	[illegible]
COLUMN TOTAL	121	96	104	11
MEDIAN	4.0	[illegible]	3.0	3.5

CHI SQUARE = 3.1 WITH N = 29 AND K = 4

SIGN TEST (BINOMIAL) SIGNIFICANCE LEVEL

COMPARISON			1-TAIL	2-TAIL
A	VS	B	0.0047	0.0094
A	VS	C	0.0320	0.0639
A	VS	D	0.1431	0.2863
B	VS	C	0.2517	0.5034
B	VS	D	0.2122	0.4244
C	VS	D	0.1917	0.3833

TABLE 5.6 Spearman Rho Values 1 - 6 Duplicated Tests 7-20 Cold versus Hot

PLUS VALUES

	1.0	.95	.9	.85	.8	.75	.7	.65	.6	.55	.5	.45	.4	.35	.3	.25	.2	.15	.1	.05	0
BC	1		1		2		1			1	1			1	1		4		2		1
KL				1	1	1	1								2						2
MN	1		2			1			2	1	2				2						2
EG	1						1			1					2						5
IJ			1	1	1											1	1				1
FH	1										2			1	1		2				1
BE											1				1		1				1
BG							1				1						1				3
CE			1						1	1			1		1						2
CG				1	1					1											2
IF	1		1								1				1		1				1
JF			1		3		1				1				2		1				1
IH	1			1	1						3				1						
JH					3						1										1
MK					2		2		1	1				1	1		2				2
ML					1						2					1	2				
NK	1		2											1		1					4
NL	1				2											1				1	2
AD			2			1	1		1		4		1		1		2				7

Duplicates where G = Gigots, L = Loin, C = cold, H = hot.

Cold versus Hot

MINUS VALUES

.05	.1	.15	.2	.25	.3	.35	.4	.45	.5	.55	.6	.65	.7	.75	.8	.85	.9	.95	1.0	Type	N
1			1		2				2	1	1	1	1		1					GC	27
	3		2						1										2	LH	16
			1		1															LC	15
					1															GH	11
	1				2						1				1				1	GC	12
					3												1			GH	12
							1		2								1			G	8
					1										1					"	8
							1				1						1			"	10
									2				1				2			"	10
			1						2					2		1				"	12
				1											1					"	12
1				1	1				1					1						"	12
					2				2						1	1			1	"	12
									1						2					L	15
			2	2					1				2		1		1			"	15
	1		1			1			2	1							1			"	16
			3		1				1		1				1	1			1	"	16
			2		2		1			1					1	1				"	28

TABLE 5.7 Sample of Forms used by ESCA 2H Students for Scoring Meat Samples Cooked to Internal Temperatures $70^{o}/75^{o}C$ (L) and $80^{o}C$ H

Scoring Samples of Meat

Appearance, flavour and texture of meats are influenced by cooking procedures. These lamb samples have been cooked simultaneously at $177^{o}C$ on the same oven shelf.

You are presented with these samples of lamb cooked to internal temperatures of $75^{o}C$ and $80^{o}C$. Score the tenderness, juiciness and flavour of each sample as indicated below. Insert each score in the appropriate box.

Score	Tenderness	Juiciness	Flavour
1	very tender	very juicy	very good
3	tender	juicy	good
5	moderately tender	moderately juicy	moderately good
7	tough	dry	poor
9	very tough	very dry	very poor

SAMPLES:

	L	H
Tenderness		
Juiciness		
Flavour		

TABLE 6.1

```
*LIST T3A

010 BARLEY * SWEDE * GIGOT     FNSC + IMAC____9.10.78
020 22
030 280,561,363,324,178,201,861,144,818,382,221,779,275,189,213
040 281,682,363,251,178,855,861,432,119,382,221,141,237,516,918
050 282,682,363,324,956,725,634,432,818,945,954,141,488,516,918
060 283,682,363,324,178,855,861,791,556,945,195,141,488,516,918
070 284,561,363,251,178,725,412,791,556,945,221,779,237,736,918
080 285,561,363,324,178,201,861,144,119,538,221,141,237,736,213
090 286,561,363,324,178,855,634,791,119,945,221,141,237,189,213
100 287,561,145,324,956,725,861,144,119,382,954,141,488,189,532
110 288,682,145,324,178,855,412,144,119,382,954,141,488,189,532
120 289,682,145,324,956,855,634,432,556,382,954,141,488,736,532
130 290,561,145,324,178,855,412,432,119,382,954,141,488,189,532
140 291,682,363,324,178,725,634,791,556,538,954,141,275,516,918
150 292,561,145,251,956,855,412,432,556,945,221,141,237,516,213
160 293,561,145,251,178,855,634,432,556,382,954,141,488,189,532
170 294,682,145,324,178,855,634,791,556,538,221,779,237,189,213
180 296,561,145,324,956,725,412,144,818,538,0,0,0,0,0
190 297,561,145,324,178,855,412,432,818,382,954,141,488,189,213
200 298,561,145,251,178,855,634,791,556,945,954,426,275,516,918
210 299,561,363,251,178,725,861,144,818,538,221,141,488,516,918
220 300,682,363,324,178,855,634,144,119,538,221,426,275,516,918
230 301,561,363,251,178,201,861,791,119,538,221,141,275,189,918
240 303,561,145,324,178,855,634,432,556,538,195,426,488,516,532
```

BARLEY * SWEDE * GIGOT FNSC + IMAC____9.10.78

RESPONSE PATTERN-----TASTING

0	1	0	1	0	1	0	1	1	0	0	0	0	1	0	1	0	1	0	0
1	0	0	1	1	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1
1	0	0	1	0	1	1	0	0	0	1	1	0	0	0	0	1	1	0	0
1	0	0	1	0	1	0	1	0	1	0	0	0	1	1	0	0	0	1	0
0	1	0	1	1	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0
0	1	0	1	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	1
0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	0	1
0	1	1	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	1
1	0	1	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	1
1	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	1	0	1	0
0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	0	1
1	0	0	1	0	1	0	1	0	0	1	1	0	0	1	0	0	0	1	0
0	1	1	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0
0	1	1	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0
1	0	1	0	0	1	0	1	0	1	0	1	0	0	1	0	0	0	1	0
0	1	1	0	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0	0
0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	1	0	0
0	1	1	0	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0
0	1	0	1	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
1	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1
0	1	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1
0	1	1	0	0	1	0	1	0	1	0	1	0	0	0	0	1	0	1	0
8	14	11	11	7	15	5	17	3	13	6	9	6	7	7	7	8	5	9	8

RESPONSE PATTERN----SNIFFING

0	1	0	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0
0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1
1	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	1
1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	1
1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0
1	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0	1	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0
0	0	1	1	0	0	0	0	1	0	0	1	1	0	0	0	0	1
1	0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	1	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0
0	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0
0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0
1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	1
0	0	1	0	0	1	0	0	1	0	1	0	1	0	0	0	0	1
0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1
0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
0	0	1	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0
6	8	8	10	2	10	4	3	15	7	10	5	10	3	9	7	6	9

TABLE 6.2

BARLEY * SWEDE * GIGOT	FNSC + IMAC____9.10.78			
280	4	1	2	1
281	2	3	2	1
282	2	2	1	1
283	3	1	1	1
284	3	1	1	1
285	4	2	2	1
286	4	1	3	0
287	2	3	2	0
288	2	1	2	0
289	1	1	1	0
290	3	2	2	0
291	3	1	1	3
292	1	1	2	0
293	2	1	2	0
294	2	0	2	1
296	2	1	0	1
297	3	1	2	0
298	2	0	0	2
299	3	2	2	2
300	3	1	1	3
301	3	2	3	3
303	3	1	0	1

PAIRED COMPARISON PREFERENCE DISTRIBUTION

	0	1	2	3	4
ACTUAL FREQ.	0	2	8	9	3
EXPECTED FREQ	1.4	5.5	8.3	5.5	1.4

CHI SQUARE = 7.76

TRIANGLE TEST TASTING

PREFERENCES	0	1	2	3	4
0	0	0	0	0	0
1	0	2	0	0	0
2	2	3	1	2	0
3	0	6	3	0	0
4	0	2	1	0	0
TOTAL	2	13	5	2	0
EXPECTED	4.3	8.7	6.5	2.2	.3

CHI SQUARE = 4.04

TRIANGLE TEST ODOUR DETECTION PRECOOKING FATS

PREFERENCES	0	1	2	3
0	0	0	0	0
1	0	1	1	0
2	2	1	5	0
3	1	4	3	1
4	0	0	2	1
TOTAL	3	6	11	2
EXPECTED	6.5	9.8	4.9	.8

CHI SQUARE = 12.72

TRIANGLE TEST ODOUR DETECTION POST-COOKING FATS

PREFERENCES	0	1	2	3
0	0	0	0	0
1	2	0	0	0
2	3	4	1	0
3	2	3	1	3
4	1	2	0	0
TOTAL	8	9	2	3
EXPECTED	6.5	9.8	4.9	.8

CHI SQUARE = 7.97

TABLE 6.3

QMC/ESCA RESEARCH PROJECTS

ASSESSMENT OF LAMB FLAVOUR

Tests may be carried out in any order but make sure that you have completed them all. Take a short break between tests if you find your senses becoming fatigued. Sips of water clear the palate.

TEST 1 - Paired Comparison (Preference) - Flavour

Decide which sample of each pair you prefer on the basis of its flavour. Record your preference by a tick in the table below. Repeat for each of the three remaining pairs.

(Retasting is allowed but too much retasting may be confusing)

1st Pair		2nd Pair		3rd Pair		4th Pair	
682		363		251		178	
561		145		324		956	

TEST 2 - Triangle Tests - Flavour

Three samples of lamb are provided. Of the three samples, two are identical and the third is different. Identify the odd sample on the basis of its flavour. Record your decision by a tick in the table below. Repeat for the second batch of three samples.

1st Batch		2nd Batch		3rd Batch		4th Batch	
201		634		791		818	
725		861		432		119	
855		412		144		556	

TEST 3 - Triangle Tests - Aroma

Each of the six beakers contains three tubes. Two tubes contain identical samples. The third contains a different (odd) sample. Identify the odd sample on the basis of its aroma. Record your decision by a tick in the table below.

(The/

(The beakers should be transferred to the assessment section. Remove the stopper from one of the tubes. Sniff contents gently. Replace stopper. Assess other two tubes).

Beaker 1		Beaker 2		Beaker 3		Beaker 4		Beaker 5		Beaker 6	
538		954		426		237		189		532	
945		195		141		488		516		918	
382		221		779		275		736		213	

COMMENTS:

NAME: DATE:

TABLE 8.1

QUEEN MARGARET COLLEGE

LAMB TRIALS 1978

To: Professor J H D Prescott

From: Mrs D A Parry

Date: 27th November 1978

Subject: Lamb Trials 1978 - Flavour Profile Analyses

7.11.78	Grass Fed Samples	388	(Loin LHS)
	Rape Fed Samples	194	(" ")
		1124	(" ")
		(Code 112)	

Report by Mrs M P Woods - Flavour Profiles of 388, 194 and 112

1. Differences do exist between the three samples. In order of quality (i.e. a combination of flavour texture and aftertaste). 194 (R), 388 (G), 112 (R). However, it is possible that untrained tasters would not be able to differentiate. This is particularly so on consideration of the fact that Flavour Profile Analysis includes an aroma assessment and texture plays a large part in influencing the profile (see note at end).

2. Sample 388 (G) demonstrates an 'alien' aroma described as cheesey/stale milk power/perfumed?

3. Sample 194 (R) demonstrates a cereal aroma.

4. Both samples 112 (R) and 388 (G) demonstrate mouthfeel effects not present in 194, e.g. mouthprickle, throat catching - these detract from aromatic detection.

5. Sample 112 (R) demonstrates a slow release of flavour volatiles due possibly to its more fibrous chewy and less soft texture.

6. In texture terms 194 (R) and 388 (G) $>>$ 122 (R) (But 388 (G) more variable).

Final/

TABLE 8.1 (Contd)

Final Observation

In spite of the cereal note in the aroma sample, 194 (R) demonstrates a more uniformly soft texture. It is more juicy and has no adverse mouthfeel effects. These facts are consistent with a product with more meat identity leaving a clean feel to the mouth.

NB Texture cannot be ignored in flavour profile assessment as the texture e.g. amount of chewing necessary affects the rate of release of volatiles and the order of appearance. As lamb is usually consumed as a hot dish some assessment of the products at a higher temperature would be desirable.

15.11.78	Grass Fed Sample	388	(Gigot LHS)	Code 625
	Rape Fed Samples	194	(" ")	Code 559
		1124	(" ")	Code 742

Report by Mrs M P Woods - Flavour Profiles of 625, 559 and 742

Once again differences between the samples were evident although they were more difficult to describe than previously. The texture dominated in that it affected release of volatiles and in this group of samples it was felt that there was variation literally from bite to bite.

The main points emerging were:

559 (R)	Late meat identity - chewy bouncy texture - aged meat?
742 (R)	Roasted meat identity
742 (R) / 625 (G)	Similar texture
625 (G)	Mouthfeel off note in flavour

Normally flavour profile analysis would be carried out after significant differences between products had been determined statistically. In a product such as meat the profile changes with each bite and 2 profiles combine, i.e. that associated with aromatics and primary tastes and that associated with texture, but the latter influences the former.

It view of the obvious variations between mouthfuls, it is felt that the/

Table 8.1 (Contd)

the differences found by profiling may only be of academic interest as all the samples would come into the acceptable category and no one difference in any product was dominent enough to pinpoint the sample completely.

It should be noted that discussion on the sessions and reports between DAP and the tasters have yet to take place.

TABLE 9.1 Sample of Forms used by ESCA 2H Students in Triangle Tests and Paired Comparisons

Triangle Test (Aroma)

Six beakers are provided. Each contains three tubes. Remove the first beaker from waterbath. Sniff each of the three tubes carefully. Assessment of aroma should be carried out carefully well away from distraction by other participants in the tests.

Of the three samples marked, two are identical, the third is different in aroma.

Please identify the odd one. Indicate your decision by a tick in the appropriate box.

Beaker 1		Beaker 2		Beaker 3		Beaker 4		Beaker 5		Beaker 6	
538		954		426		237		189		532	
945		195		141		488		516		918	
382		221		779		275		736		213	

Triangle Test (Flavour)

Of the three meat samples, two are identical and the third is different. Identify the odd sample on the basis of its flavour. Indicate your answer by a tick in the appropriate box.

201	
725	
855	

TABLE 9.1 (Contd)

Paired Comparison Test (Preference)

You are presented with a pair of samples 45 and 62. Please indicate by a tick in the appropriate box, the sample which you prefer in relation to its flavour.

45	
62	

Paired Comparison Test (Preference)

You are presented with a pair of samples 613 and 437. Please indicate by a tick in the approriate box, the sample which you prefer in relation to its general acceptability.

613	
437	

TABLE 9.2 Grass and Lucerne Silages

1. Composition of Grass and Lucerne Silages

	Grass Silage		Lucerne Silage	
	Batch 1 (Up to 21.11.78)	Batch 2 (From 22.11.78)	Batch 1 (Up to 21.11.78)	Batch 2 (From 22.11.78)
Toluene DM (gKg^{-1})	241	233	274	264
pH	4.61	4.21	4.07	4.09
Composition of DM ($gKg^{-1}DM$)				
OM	912	918	926	925
TN	31.8	29.4	21.4	21.9
PN	430	513	451	564
NH_3N	109	43	51	48
WSC	6	26	100	98
MAD - fibre	355	355	313	320
Gross energy ($MJKg^{-1}DM$)	18.5	18.7	18.4	18.7

where DM = dry matter OM = organic matter TN = total nitrogen PN = protein nitrogen

WSC = water soluble carbohydrate MAD - fibre = modified acid detergent fibre

Both silages were prepared by the addition of formic acid/formaldehyde. Ethanol, formic acid, acetic acid, propionic acid, n-butyric acid and lactic acid were also present in both silages.

2./

TABLE 9.2 (Contd)

2. Feed Efficiency of the Diets (gLWG/KgDM intake)

	Grass	Lucerne
Silage only	-4	30
Silage + barley (from 22.11.78)	124	92

3. Digestibility Coefficients, Digestible Nutrients, Nitrogen Retention and Metabolisable Energies

	Digestibility	Digestible Crude Protein ($gKg^{-1}DM$)	Digestible Energy ($MJKg^{-1}DM$)	Nitrogen Retention (g/day^{-1})	Metabolisable Energy ($MJKg^{-1}DM$)
Grass Silage	0.712	109.70	13.42	7.09	11.14
Grass Silage + barley	0.720	107.10	13.40	8.22	11.23
Lucerne Silage	0.610	159.70	10.86	5.79	8.44
Lucerne Silage + barley	0.637	149.30	11.24	3.79	8.75

4. Mean Intakes of Silage and Silage + Barley and Live Weight Gains Over 28 Days

	Intake ($gKg^{-1}LW$)	Live Weight Gains (g/day^{-1})
Grass Silage	25.8	-0.4
Grass Silage + barley	22.3	92
Lucerne Silage	32.9	46
Lucerne Silage + barley	27.8	96

TABLE 9.3

Details of Forage Crops

1. The Design of the Experiment for the Production of Store Lambs 1977-78.

System	Oct -Dec	Dec -Mar
1	Rape	Swedes
2	"	Cabbages
3	"	Yellow Turnips
4	"	Barley (Indoors)
5	Dutch Turnip	Swedes
6	"	Cabbages
7	"	Yellow Turnips
8	"	Barley (Indoors)
9	Grass aftermaths	Swedes
10	"	Cabbages
11	"	Yellow Turnips
12	"	Barley (Indoors)

A small supplement of whole barley (135g) was introduced on 24.11.77. For groups grazing swedes, cabbages and yellow turnips this supplement was gradually increased as indicated in Chapter 9.

2. Details of the Crops 1977

The rape (Brassica napus cv. Lair) and stubble (Dutch) turnips (Brassica campestris cv. Civasto) were sown as catch crops following a grass crop (which had been converted to silage) during July. As noted above, they were grazed by groups along with grass aftermaths from October to December. Total crop dry matter (DM) yield was 20-30% higher from the Dutch white turnips.

The main crops were winter Swedes (Brassica napa cv. Wilhelmsburger and Doon Major), Cabbages (Brassica oleracea cv. Celtic and Langendijk 3) and Yellow Turnips (Brassica rapa cv. Aberdeen yellows) which were grazed from December to March. A fourth group was also fed/

fed whole barley indoors at this time. Stocking rate was adjusted with a fixed DM allowance per animal for the expected period of grazing (0.035 Kg DM per Kg liveweight per day). Strip grazing was practised.

At the same time, lucerne established in 1976 was harvested three times in 1977. The first two cuts were used for comparison with grass silage involving metabolism studies with wether lambs. Details of these two silages are shown in Appendix Table 9.2. Palatability studies were carried out on wethers from these two regimes during 1978 when a second comparison of wether lambs fed the same variety of rape (Brassica napa cv. Lair) with grass fed lambs was also made. Details of the barley supplementation of the silage fed lambs are given in Chapter 9. No supplement was given to the grass and rape fed wethers preslaughter.

3. Mean Weights of Lambs as Sold (Kg)

Rape	36.1
Dutch Turnips	38.9
Aftermaths	36.6
Swedes	41.7
Cabbages	41.3
Yellow Turnips	41.3
Barley	42.0

4. Live Weight Gains (g/day)

Rape	110
Dutch Turnips	151
Aftermaths	82
Swedes	74
Cabbages	71
Yellow Turnips	68
Barley	83*

*Lambs lost weight over initial adjust period from forages onto barley. From January 17th to 7th March 1978 growth rate became 165g/day.

TABLE 9.4

Carcass Weights of Lambs Used in the Trials

Feeding Regime	Carcass No.	Weight (kg)	Slaughter Date
Grass	388	18.5	6.12.77
"	1240	17.5	"
"	893	16.5	"
"	914	16.0	"
Rape	902	16.5	"
"	1124	18.5	"
"	1464	15.5	"
"	194	18.5	"
Stubble Turnips	1361	18.0	"
"	384	17.0	"
"	436	15.5	"
"	1337/1366	19.5	"
Turnip	1458	20.0	12.2.78
"	1462	20.0	"
"	1165	16.5	"
"	1255	18.0	"
Barley	54	22.5	12.2.78
"	1548	19.0	"
"	454	16.5	"
"	107	20.0	"
Cabbages	941/970	23.0	4.3.78
"	1150	23.0	"
"	1160	20.5	"
"	586	21.5	"
Swedes	550	25.0	4.3.78
"	364	21.0	"
"	365	22.5	"
"	888	25.5	"
Whole Barley*	1060	17.25	4.5.78
"	651	16.25	"
"	970	19.5	"
"	390	18.0	"
Grass**/			

Grass**	555	18.6	20.11.78
"	1310	18.3	"
"	787	18.6	"
"	569	17.7	"
"	90	17.2	"
"	975	18.0	"
Rape**	W153	18.4	"
"	W138	18.3	"
"	382	17.7	"
"	1468	18.3	"
Grass Silage**	466	16.9	8.1.79
"	974	18.7	"
"	95	16.3	"
"	447	17.7	"
Lucerne Silage**	325	19.4	"
"	361	16.9	"
"	48	18.6	"
	469	17.9	"

* Suffolk x Dorset/Finn females. Used in trials described in Chapter 7.

** Estimated on basis of 46-47% of slaughter weight. (J. FitzSimons recommendation for lambs of this type). These lambs were sold to FMC. Only the joints required were weighed.

BIBLIOGRAPHY

Ackerson B.A., Johnson R.R. and Hendrickson R.L. (1976)
Effects of treatment of whole fat soybeans or soy flour with formaldehyde to protect the polyunsaturated fatty acids from biohydrogenation in the rumen.
J.Nut. 106, 1383-1390

AHEA Terminology Committee (1971)
Handbook of Food Preparation, 6th Edition (Rev.)
Amer.Home Econ. Assoc., Washington D.C.

American Society for Testing and Materials (1968)
Correlation of Subjective-Objective Methods in the Study of Odours and Taste.
ASTM.

AMSA (1978)
Guidelines for Cookery and Sensory Evaluation of Meat.
Am. Meat Science Assoc., National Live Stock and Meat Board, Chicago, Illinois.

Amerine M.A., Pangborn R.M. and Roessler E.B. (1965)
Principles of the Sensory Evaluation of Food
Academic Press.

Batcher O.M., Dawson E.H., Pointer M.T. and Gilpin G.L. (1962)
Quality of Raw and Cooked Lamb Meat as Related to Fatness and Age of Animal.
Food Tech. 16, 102-104 107-111.

Batcher O.M., Brant A.W. and Kunse M.S. (1969)
Sensory evaluation of lamb and yearling mutton flavour.
J.Food Sci. 34, 272-274.

Bennett C.A. and Franklin N.L. (1954)
Statistical Analysis in Chemistry and the Chemical Industry.
Chapters 10 and 11.
Wiley, New York.

Biometrika Tables for Statisticians (1972, reprinted with corrections 1976)
Pearson E.S. and Hartley H.O. (Eds.)
Cambridge University Press.

Birch G.G., Brennan J.G. and Parker K.J. (1977)
Sensory Properties of Foods.
Applied Science Publishers, London.

Blakesley C. Newton. (1977)
An aid to statistical evaluation of organoleptic panel results.
Lebbensmittel-Wissenschaft und Technologie 10, (1) 21-23.

Bock R.D. and Jones L.V. (1968)
The measurement and prediction of judgement and choice. Ch.6.
Holden Day, San Francisco.

Boggs M.M. and Hanson H.L. (1949)
Analysis of Foods by Sensory Difference Tests.
Advances in Food Research, 2, 219-258

Bouton P.E., Harris P.V. and Shorthose W.R. (1971)
Effect of ultimate pH upon the waterholding capacity and tenderness of mutton.
J.Food Sci., 36, 435-439

Bouton P.E., Harris P.V. and Shorthose W.R. (1972a)
The effects of ultimate pH on ovine muscle: water holding capacity.
J.Food Sci. 37, 351-355.

Bouton P.E., Harris P.V. and Shorthose W.R. (1972b)
The effects of ultimate pH on ovine muscle: mechanical properties.
J.Food Sci., 37 356-360

Bruvold W.H. (1978)
Laboratory panel estimation of consumer assessments of taste and flavour.
J.Applied Psychology 54, 326-330

Buttery R.G., Ling L.C., Teranishi R. and Mon R.T. (1977)
Roasted lamb fat: basic volatile components
J.Agric.Food Chem., 25, 1227-1229.

BS 5098 (1975)
Glossary of Terms Relating to the Sensory Evaluation of Food

BS 5929 (1980)
Methods for Sensory Analysis of Food. Part 1. Introduction and general guide to methodology.

Byer A.J. (1964)
Looking askance at statistical sensory testing.
Food Tech., 18 (11), 59-64.

Cagan R.H. and KareM.R. (1981)
Biochemistry of Taste and Olfaction
Academic Press.

Cain W.S. (1982)
Odour identifications by males and females: predictions vs performance.
Chemical Senses, 2, 129-142

Campbell A.M., Penfield M.P. and Griswold R.M. (1979)
The Experimental Study of Food
Houghton-Mifflin/Constable.

Carpenter Z.L., Smith G.C., King G.T. and Hoke K.E. (1969)
Lamb carcass maturity and its relationship to palatability.
J.Anim.Sci., 30, 496-502

Carpenter Z.L., Smith G.C., King G.T. and Hoke K.E. (1970)
Palatability of yearling mutton carcasses.
J.Anim.Sci., 31, 310-317.

Caul J.F. (1957)
The profile method of flavour analysis.
Advances in Food Research, 7, 1-40.

Caporaso F., Sink J.D., Dimick P.S., Mussinan C.J. and Sanderson A. (1977)
Volatile Components of Ovine Adipose Tissue.
J. Agric. Food Chem., 25, 1230-1233

Chang S.S. and Petersen R.J. (1977)
Recent developments in the flavor of meat.
J.Food Sci., 42, 298-305.

Christie E. (1962)
Conduct of tasting tests in Australia.
Food Tech., 14, 77, 124-125, 161-162, 165, 169 and 171.

Civille G.V. (1978)
Case studies demonstrating the role of sensory evaluation in food product development.
Food Tech., 32, 59-60.

Cole D.J.A. and Lawrie R.A. (Eds.) (1975)
Meat
Butterworths.

Corbett J.L., Furnival E.P., Southcott W.H., Park R.J. and Shorthose W.R. (1973)
Induced cryptorchism in lambs. Effects of growth rate, carcass and meat characteristics.
Anim. Prod., 16 (2), 157-163.

Cover S. (1959)
Meat cookery from the scientific viewpoint.
Proc. Res. Conf. Am. Meat Inst. Found., 11, 99-111.

Cornford S.J. (1977)
Sensory methods in the milling and baking industries.
Proceedings of a Joint Symposium 'Sensory Quality Control'
IFST/SCI January 1977, 60-68.

Cramer D.A., Barton R.A., Shorland F.B. and Czochanoka Z. (1967)
A comparison of the effects of white clover (Trifolium repens) and of perennial rye grass (Lolium perenne) on fat composition and flavour of lamb.
J.Agric.Sci., Cambridge, 367-373

Cramer D.A., Pruett J.B., Swanson V.B., Schwartz W.C., Kattnig R.M. Phillips B.L. and Wookey L.E. (1970)
Comparing Breeds of Sheep. J.Anim.Sci. 30, 1031(A)
and as reported by Ford and Park (1980)

Crocker E.C. (1948)
Flavor of meat.
Food Res., 13, 170-183.

Cross H.R., Smith G.C. and Carpenter Z.L. (1970)
Relationship of U.S.A. Dept. Carcass Grades to palatability of lamb cuts.
Tex. Agr. Exp. Sta. Prog. Rep. 2754, 29-31.

Cross H.R., Moen R and Stanfield M.S. (1978)
Training and testing of judges for sensory analysis of meat quality.
Food Tech., 32 48-54.

Cross H.R., Stanfield M.S., Elder R.S. and Smith G.S. (1979)
A comparison of roasting versus broiling on the sensory characteristics of beef longissimus steaks.
J.Food Sci., 44, 310-313.

Czochanska Z., Shorland F.B., Barton R.A. and Rae A.L. (1970)
A note on the effect of the length of the resting period before slaughter on the intensity and flavour of lamb.
N.Z.J.Agric., 13, 662-663.

Dransfield E., Nute G.R.MacDougall D.B. and Rhodes D.N. (1979)
Effect of Sire Breed on Eating Quality of Cross-bred Lambs.
J.Sci.Food Agric., 30, 805-808.

Eating for Health (1978)
H.M.S.O.

Ellis B.H. (1968)
Preference Testing Methodology, Part I.
Food Tech., 22, (5), 583-588, 590.

Erhart J.P. (1978)
The role of the sensory analyst in product development.
J.Food Tech., 32, 57-58, 66.

Ford A.L., Park R.J. and Shorthose W.R. (1972)
Unpublished data reported in Chapter 9.
Developments in Meat Science. (Ed.) Ralston Lawrie
Academic Press (1980).

Ford A.L., Park R.J. and McBride R.J. (1975)
Effect of a protected lipid supplement on flavour properties of sheep meats.
J.Food Sci., 40, 236-239.

Ford A.L. and Park R.J. (1980)
Developments in Meat Science. (Ed.) Lawrie R.A.
Academic Press
219-248.

Forss D.A. (1969)
Role of lipids in flavor.
J.Agric.Food Chem., 17, 681-685

Francis F.J. and Clydesdale F.M. (1975)
Food Colorimetry - Theory and Application.
The AVI Publishing Co., Connecticut.

Francis F.J. (1977)
Colour and Appearance as Dominating Properties of Foods.
Birch C.G., Brennan J.G. and Parker K.J.
Sensory Properties of Foods. 27-45.
Applied Science Publishers Ltd, London.

Freeman G.C. and Mossadeghi N. (1972)
Influence of sulphate nutrition on flavour components of three cruciferous plants, radish (Raphanus sativas), cabbage (Brassica olevacea capitata), and white mustard (Sinapis alba).
J.Sci. Food Agric., 23, 387-402.

Friedman M. (1937)
The use of ranks to avoid the assumption of normality implicit in the analysis of variance.
J.Amer.Statist. Assoc., 32, 675-701.

Friedman M. (1940)
A comparison of alternative tests of significance for the problem of m rankings.
Ann. Math. Statist., 11, 86-92

Frijters J.E.R. (1977)
The effect of duration of intervals between olfactory stimuli in the triangular method.
Chemical Senses and Flavour, 2, 301-311.

Frijters J.E.R. and Beumer-Stoffer S.C.C.(1978)
Comparison of storage time - termperature effects on sensory and hedonic attributes of frozen and deep-frozen chickens.
Br. Poult. Sci., 19, 225-232.

Frijters J.E.R. (1979)
The paradox of discriminatory and nondiscriminators resolved.
Chemical Senses and Flavour, 4, 355-358.

Garrett W.N., Young Y.T., Dunkley W.L. and Smith L.M. (1976)
Increasing the polyunsaturated fat content of beef and lamb.
J.Anim. Sci., 42, 845-853.

Girardot N.F., Peryam D.R. and Shapiro R. (1952)
Selection of sensory testing panels.
Food Tech., 6, 140-143

Goodman L.A. (1954)
Kolmogorov-Smirnov tests for psychological research.
Psychol. Bull., 51, 160-168.

Gormley T.R. and Sherington J. (1978)
Assessment of taste panellists.
Irish J. Food Sci and Tech., 2, 59-66.

Greenhalgh C. (1970)
Discrimination testing: further results and developments.
Esomar Congress, Barcelona 1971 - 181-199.
(Market Research Abs. 14, 1970, Ref: 1239)

Gregson R.A.M. (1960)
Bias in the measurement of food preferences by triangular tests.
Occupational Psychol., 34, 249-257.

Gridgeman N.T. (1955)
Taste comparisons: Two samples or three?
Food Tech., 9, 148-150.

Gridgeman N.T. (1970)
A re-examination of the two-stage triangle test for the perception of sensory differences.
J.Food Sci., 35, 87-91.

Griffiths N.M. and Patterson R.L.S. (1970)
Human olfactory responses to 5α-androst-16-en-3-one principal conponent of boar taint.
J.Sci. Food Agric., 21, 4-6.

Griswold R.M. (1962)
The Experimental Study of Food
Constable, London.

Haber A. and Runyon R.P. (1977)
General Statistics (3rd Ed.)
Addison-Wesley, London.

Harper R., Bate-Smith E.C., Land D.G. and Griffiths N.W. (1968a)
A glossary of odour stimuli and their qualities.
Perf. Essen. Oil Res., 59, 22-37

Harper R., Bate-Smith E.C. and Land D.G. (1968b)
Odour description and classification
Churchill.

Harper R (1972)
Human Senses in Action
Churchill Livingstone.

Harper R. (1977)
A short history of sensory analysis in the United Kingdom.
Sensory Properties of Foods. 167-187
(Eds.) Birch G.G., Brennan J.G. and Parker K.J.
Applied Science Publishers.

Harries J.M. (1955)
Positional bias in sensory tests.
Food Tech., 9, 86-90.

Harries J.M., Robertson J. and Walmsley R. (1960)
Weight loss of beef during roasting
Home Economics, Sept. 1960 (Reprint supplied by principal author)

Harries J.M., Bryce Jones K., Houston T.W. and Robertson J. (1963)
Studies in Beef Quality: Development of a system for assessing palatability.
J.Sci.Food Agric., 14, 501-509.

Harries J.M. (1973)
Complex Sensory Assessment.
J.Sci.Food Agric., 24, 1571-1581.

Harrison S. and Elder L.W. (1950)
Some applications of statistics to laboratory taste testing.
Food Tech., 4, 434-439

Harries J.M. and Smith G.L. (1982)
The two-factor triangle test.
J.Food Tech., 17, 153-162

Headley M.E. and Jacobson M. (1960)
Harrison S. and Elder L.W. (1950)
Some applications of statistics to laboratory taste testing.
Food Tech., 4, 434-439.

Headley M.E. and Jacobson M. (1960)
Electronic and conventional cookery of lamb roasts.
J. Amer. Dietetic Assoc., 36, 337-340.

Hollander M. and Wolfe D.A. (1973)
Nonparametric Statistical Methods.
John Wiley and Sons, Inc.

Hornstein I. and Crowe P.F. (1964)
Meat flavour: a review.
J Gas Chromatogr., 3, 128-131.

Hornstein I and Crowe P.F. (1963a)
Meat Flavour - A review.
J. Gas Chromatography, 11, 128-131.

Hornstein I. and Crowe P.F. (1963b)
Meat Flavour: Lamb.
J. Agric. Food Chem, 11, 147-149

Hotelling H. and Pabst R. (1936)
Rank correlation and tests of significance involving no assumptions of normality.
Int. Encycl. Unif. Sci., 2, No.7, (Univ. of Chicago Press).

Household Food Consumption and Expenditure: 1979.
Annual Report of the National Food Survey Committee. H.M.S.O.

Howgate P. (1977)
Measurement of fish by an objective sensory method.
Proceedings of a Joint Symposium 'Sensory Quality Control'
IFST/SCI January 1977. 41-48.

Howgate P and Smith G.L. (1981)
Processing and interpeting hedonic rating data.
International Symposium 'The Measurement of Acceptability'
25.3.81 SCI Food Group with the Royal Statistical Society.

Hutchings J.B. (1977)
The importance of the visual appearance of foods to the food processor and the consumer. 45-56.
The Sensory Properties of Foods (Eds.) Birch A.G., Brennan J.G. and Parker K.J. Applied Science Publishers, London.

Jacobs J.A., Field R.A., Botkin M.P., Kaltenbad C.C. and Riley M.L. (1972)
Effects of testerosterone ethanate on lamb carcass composition and quality.
J.Anim.Sci., 34, 30-36.

Jagusch K.T. (1975)
Production of lamb with polyunsaturated depot fats (from feeding a dietary supplement of formaldehyde-treated ground sunflower seed).
N.Z.J.Agric.Res., 18, 9-12.

Jellinek G. (1964)
Introduction to and critical review of modern methods of sensory analysis (odour, taste and flavour evaluation) with special emphasis on descriptive sensory analysis (flavour profile method).
Journal of Nutrition and Dietetics, Vol.1., 219-260.

Jeremiah L.E., Smith G.C. and Carpenter Z.L. (1972) Ovine Yields II: Palatability attributes within various quality grades.
J.Anim.Sci., 34, 196-202.

Jones L.V., Peryam D.R. and Thurstone L.L. (1955)
Development of a scale for measuring solders' food preferences.
Food Research, 20, 512-520.

Kare M.R. and Lammer O. (Eds.) (1977)
The Chemical Senses and Taste.
Academic Press.

Kemp J.D., Shelley J.M., Ely D.G. and Moods W.G. (1972)
Effects of castration and slaughter weight on fatness, cooking losses and palatability of lamb.
J.Anim.Sci., 34, 560-562.

Kemp J.D., Johnson A.E., Steward D.F., Ely D.G. and Fox J.D. (1976)
Effect of dietary protein, slaughter weight and sex on carcass composition, organoleptic properties and cooking losses of lamb.
J.Anim.Sci., 42 (3), 575-583

Kendal M.G. and Stuart A. (1979)
Advanced Theory of Statistics, Volume 2 (3rd Edition) 468-477.
Griffin, London.

King B., Wyler R and Solms J. (1978)
Problems of Flavour Application in Food Systems:in Progress in Flavour Research 327-335.
(Eds.) Land D.G. and Nusten H.E.
Applied Science Publishers.

Kirton A.H. and Pickering F.S. (1967)
Factors associated with differences in carcass conformation in lamb.
N.Z.J.Agric.Res., 10, 183-200.

Kramer A (1960)
A rapid method for determining significance of differences from rank sums.
Food Tech., 14, 576-581.

Kramer A., Murphy A.M., Briant A.M., Wang M. and Kirkpatrick M.E. (1961)
Studies in tastepanel methodology.
J.Agric.Food Chem., 9, 224-228.

Kramer A., Kahan G., Cooper D and Papavasiliou A. (1974)
A non-parametric ranking method for the statistical evaluation of sensory data.
Chem. Senses and Flavour, 1, 121-133.

Jackson T.H. (1968)
(Personal communication)

Land D.G. and Nursten H.E. (Eds.) (1978)
Progress in Flavour Research.
Applied Science Publishers, London.

Land D.G. and Piggot J. (1981)
Assessing variations in liking of foods.
International Symposium organised jointly by the Sensory Panel Society of Chemical Industry and the Royal Statistical Society, 25th March 1981.

Larmond E. (1970)
Methods for the sensory evaluation of food.
Canadian Department of Agriculture, publication no. 1284.

Lawrie R.A. (1966)
Meat Science (1st Edition)
Pergamon Press.

Lawrie R.A. (1979)
Meat Science (3rd Edition)
Pergamon Press

Lawrie R.A. (Ed.) (1980)
Developments in Meat Science, Vol.1.
Academic Press.

Le Magnen L., MacLeod P. (1977)
Olfaction and Taste VI.
Academic Press.

Lockhart E.E. (1951)
Binomial systems and organoleptic analysis.
Food Tech., 5, 428-431.

Mackey A.O. and Jones P. (1954)
Selection of members of food tasting panels: discernment of primary tastes in water solution compared with judging ability for foods.
Food Tech., 8, 527-530.

Macy R.L., Naumann H.D. and Bailey M.E. (1963a)
Water soluble flavor and odor precursors of meat, 1.
J.Food Sci., 29, 136-141.

Macy R.L., Naumann H.D. and Bailey M.E. (1963b)
Water soluble flavor and odor precursors of meat, 2.
J.Food Sci., 29, 142-148.

McBride R.L. and Laing D.G. (1979)
Threshold determination by triangle testing: effects of judgemental procedure, positional bias and incidental training.
Chemical Senses and Flavour, 4, 319-326.

Meyer L.H. (1960)
Food Chemistry
Reinhold Publishing Corporation, New York.

Miller G.A. (1956)
The magic number seven, plus or minus two: some limits on our capacity for processing information.
Psychol Rev., 63, 81-97.

Misock J.P., Campion D.R., Field R.A. and Riley M.L. (1972)
Palatability of heavy ram lambs.
J.Anim.Sci., 34, 1440-1444.

Moncrieff R.W. (1951)
The Chemical Senses (2nd Edition)
Leonard Hill, London.

Moncrieff R.W. (1970)
Odours.
Wm. Heinemann Medical Books Ltd., London.

Morgan J.H.L. (1972)
Effects of low plane of nutrition on growth and eating quality of steers.
J.Agric.Sci., 78, 417-423.

Moskowitz H.R. (1974)
Sensory Evaluation by Magnitude Estimation.
Food Tech., 28 (11), 16-21.

Mottram D.S. (1981)
Cooking meat for sensory and instrumental assessment.
J.Sci. Food Agric., 32, 523-524.

Mottram D.S. and Edwards R.A. (1983)
The Role of Triglycerides and Phospholipids in the Aroma of Cooked Beef.
J.Sci.Food Agric., 34, 517-522.

Muller H.G. (1977)
Sensory Quality Control: Report on a Survey. Proceedings - Joint Symposium SCI/IFST Sensory Quality Control, University of Aston. 28-40.

Murray K.E., Park R.J. and Stanley G. (1976)
Unpublished data reported in Chapter 9, Developments in Meat Science, (Ed.) Ralston Lawrie.
Academic Press (1980)

Nicol A.M. and Jagusch K.T. (1971)
The effects of different types of pasture on the organoleptic qualities of lamb.
J.Sci. Food Agric., 22, 464-466.

Neyman J. (1950)
First Course in Probability and Statistics.
273, 277-278 and 350
Henry Holt, New York.

Nursten H.E. (1977)
The important flavour volatiles in foods
in Sensory Properties of Foods. 152-166.
(Eds.) Birch G.C., Brennan J.G. and Parker K.J.
Applied Science Publishers.

Oliver W.M., Carpenter Z.L., King G.T. and Shelton J.M. (1967)
Qualitative and quantitative characteristics of ram, wether and ewe lamb carcasses.
J.Anim.Sci., 26, 307-310.

Park R.J., Corbett J.L., Furnival E.P. (1972a)
Flavour differences in lambs grazed on lucerne (Medicago sativa) or phalaris (Phalaris tuberosa) pastures.
J. Agric. Sci., Camb. 78, 42-52.

Park R.J., Spurway R.A. and Wheeler J.L. (1972b)
Flavour differences in meat from sheep grazed on pasture or winter forage crops.
J. Agric.Sci., Camb. 78, 53-56.

Park R.J. and Minson D.J. (1972)
Flavor differences in meat from lambs grazed on tropical legumes.
J. Agric. Sci., 79 (3), 473-475.

Park R.J., Murray K.E. and Stanley D. (1974)
4-hydroxydodec-cis-6-enoic acid lactone: An important component of lamb flavour from lambs fed a lipid-protected dietary supplement.
Chem.Ind., London. 380-384.

Park R.J. and Ford A.L. (1975)
Effect on meat flavour of period of feeding protected lipid supplement to lambs.
J.Food Sci., 40, 1217-1221

Park R.J., Ford A., Minson D and Baxter R.I. (1975)
Lucerne-derived flavour in sheep meat as affected by season and duration of grazing.
J.Agric.Sci., 84, 209-213.

Park R.J., Ford A.L. and Ratcliff D. (1976)
The influence of two kinds of protected lipid supplements on the flavour of lamb.
J. Food Sci., 41, 633-635.

Park R.J., Ford A.L. and Ratcliff D. (1978a)
Use of protected lipid supplement to modify the flavour of mutton.
J. Food Sci., 43, 874-877.

Park R.J., Ford A.L. and Ratcliff D. (1978b)
A study of "sweet" flavour in lambs produced by feeding protected sunflower seed.
J. Food Sci., 43, 1363-1367.

Parry D.A. (1970)
The organoleptic qualities of protein foods (with special to cooking procedures in Proteins as Human Food
(Ed.) Lawrie R.A.
Butterworth and Co.
Proceedings of the 16th Easter School in Agricultural Science, University of Nottingham 1969.

Patterson R.L.S. (1975)
The Flavour of Meat in Meat. 359-379.
(Eds.) D.J.A.Cole and Lawrie R.A.
Butterworths.

Paul A.A. and Southgate D.A.T. (1978)
McCance and Widdowson's The Composition of Foods (4th Edition)
H.M.S.O.

Paul P.C., Torten J. and Spurlock G.M. (1964a)
Eating Quality of Lamb I. Effects of age and food.
J. Food Tech., 18, 1779-1782.

Paul P.C., Torten J. and Spurlock G.M. (1964b)
Eating Quality of Lamb II
Effect of preslaughter nutrition.
Food Tech., 18, 1783-1784.

Paul P.C., Torten J. and Spurlock G.M. (1964c)
Eating Quality of Lamb III
Overall comparisons and inter-relationships.
Food Tech, 18, 1785-1789.

Paul P.C. and Palmer H.H. (1972)
Food Theory and Applications
John Wiley and Sons.

Pearson A.M., Wenham L.M., Carse W.A., McLeod K., Davey C.L. and Kirton A.H. (1973)
Observations on the contribution of fat and lean to the aroma of cooked beef and lamb.
J.Anim.Sci., 36, 511-515.

Pepper F.H. and Pearson A.M. (1971)
Possible role of adipose tissue in meat flavour - the non-dialysable aqueous extract.
J. Agric. Food Chem., 19, 964-968.

Peryam D.R. and Schwartz V.W. (1950)
Measurement of Sensory Differences.
Food Tech., 4, 390-395.

Peryam D.R. and Pilgrim F.J. (1957)
Hedonic scale method of measuring food preferences.
Food Tech., 11 (9), 9-14.

Piggott J. and Land D.G. (1981)
Assessing variation in liking of foods.
International Symposium 'The Measurement of Acceptability'
25.3.81. SCI Food Group with the Royal Statistical Society.

Pratt J.W. and Gibbon J.D. (1981)
Concepts of nonparametric theory.
Springer Series in Statistics.
Springer-Verlag, Heidelberg/New York.

Price L.G. and Greene B.E. (1978)
Factors affecting panellists' perceptions of cured meat flavour.
J.Food Sci., 43, 319-336

Pridmore W.A. (1979)
Statistical analysis of sensory difference tests.
J.Sci.Food Agric., 30, 218.

Prodfact 1982.
British Farm Produce Council.

Purchas R. (1975)
Effects of feeding formaldehyde treated oil seeds to sheep on pasture.
N.Z.J.Exp. Agric., 3, 219-222.

Reineccius G.A. (1979)
Off-flavours in meat and fish - a review.
J. Food Sci., 44, 12-21.

Rhodes D.N. (1969)
Meat Res. Inst. Report H 942.

Rhodes D.N. (1969)
Meat Res. Inst. Ann. Report 1968-69
Agric. Research Council, London.

Rhodes D.N. (1970)
Meat Res. Institute Report No H942 and
in Meat Production from Entire Male Animals, 189-198.
J and A Churchill.

Rhodes D.N. (1971)
Meat Res. Institute Ann Report 1970-71
Agric. Research Council, London.

Rhodes D.N. (1971)
A comparison of the quality of meat from lambs reared intensively indoors and conventionally on grass.
J.Sci. Food Agric., 22, 667-669.

Rhodes D.N. (1973)
The Chemistry of Meat Flavour.
MRI Research Memorandum No.24.

Rhodes D.N. (1971a)
Meat flavour and consumer acceptability in
Progress in Flavour Research (Eds.) Land D.G. and Nursten H.E.
Applied Science Publishers. 307-319.

Rhodes D.N. (1971b)
A Comparison of the Quality of Meat from Lambs Reared Intensively Indoors and Conventionally on Grass.
J.Sci. Food Agric., 22, 667-669.

Roessler E.B, Pangborn R.M., Sidel J.L. and Stone H (1978)
Expanded statistical tables for estimating significance in paired-preference, paired-difference, duo-trio and triangle tests.
J. Food Sci., 43, 940-947.

Shallenberger R.S. (1971)
The Theory of Sweetness
in Sweeteness and Sweeteners
(Eds.) Brich C.G., Green L.F. and Coulson C.B. 42-50.

Shephard D. (1954)
The adequacy of everyday quantitative expressions as measurements.
Brit. J. Psychol., 45, 40-50.

Shephard D. (1955)
Descriptive terms and points systems for rating food qualities.
Food Res., 20, 114-117

Shewan J.M., MacIntosh R.G., Tucker C.G. and Ehrenberg A.S.C. (1953)
Scoring system for freshness based on odour of raw white round fish.
J.Sci. Food Agric., 4, 283-298.

Shorland F.B., Czochanska Z., Barton R.A. and Rae A.L. (1970)
Influence of pasture species on the keeping quality of lamb and mutton.
J. Sci. Food Agric., 21, 1-4.

Sidel J.L., Woolsey A and Stone H. (1976)
Experimental design and analysis of sensory tests.
Food Tech., 30 (11), 32-38

Siegel S. (1956)
Nonparametric statistics for the behavioural sciences.
McGraw Hill.

Sink J.D. (1973)
Lipid soluble components of meat flavor and odors.
J. Am. Oil Chem. Soc., 50, 82A.

Sink J.D. and Caparoso F. (1977)
Lamb and mutton flavour: contributing factors and chemical aspects.
Meat Science, 1, 119-127.

Smith G.C., Carpenter Z.L. and King G.T. (1970)
Palatability characteristics of yearling mutton carcasses.
Tex. Agric. Exp. Stat. Progr. Rep. 2754, 33-35.

Smith G.C., Carpenter Z.L. and King G.T. (1970)
Palatability characteristics of yearling mutton carcasses.
Tex. Agr. Exp. Stat. Prog. Rep. 2754, 33-35
and J. Anim. Sci., 30, 496-502.
Lamb Carcass Quality 1.Palatability of Leg Roasts.

Smith G.C., Carpenter Z.L., King G.T. and Hoke K.E. (1970)
Lamb carcass quality 1: Palatability of leg roasts.
J. Anim.Sci., 30, 496-502
Lamb carcass quality 11: Palatability of rib, loin and sirloin chops.
J.Anim.Sci.

Smith G.C. and Carpenter Z.L. (1970)
Lamb Carcass quality 111: Chemical, physical and histological measurements.
J Anim. Sci., 31, 697-706.

Snedecor G.W. and Cochran W.G. (1967)
Statistical Methods
Iowa State University Press, U.S.A.

Spencer H.W. (1972)
Sensory Measurement of Flavour
J. Food Tech. 5, (4), 198-205.

Spencer H.W. (1979)
Alternative Sensory Test Designs.
J. Sci. Food Agric., 30, 218.

Sprague E.C. and Grindlay H.S. (1907)
A precise method of roasting beef.
Univ. Illinois, Univ. Studies, Vol.11, No.4.

Steedman C.D., Hawrysh Z.J., Hardin R.T. and Roblcz A.R.(1979a)
Influence of rape seed meal on the eating quality of chicken I: Subjective evaluation by a trained taste panel and subjective measurements.
Poultry Sci., 58, 148-158.

Steedman C.D., Hawrysh Z.J., Hardin R.T. and Roblcz A.R. (1979b)
Influence of rape seed meal on the eating quality of chicken II: Consumer trials.
Poultry Sci., 58, 337-340.

Stevens M.A. (1970)
Use of Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables.
J. Royal Statist. Soc., Series B, 32, 115-122.

Stevens M.A. (1974)
Statistics for goodness of fit and some comparisons.
J. Amer. Statist. Assoc., 69, 730-737.

Stone H., Sidel J., Oliver S., Woolsey A and Singleton R.C. (1974)
Quantitative descriptive analysis.
Food Tech., 28,(11), 24-34.

Stone H. and Sidel J.L. (1978)
Computing exact probabilities in sensory descrimination tests.
J. Food Sci., 43

Teranishi R (1971)
Flavour Research, Principles and Techniques
Marcel Dekker Inc., New York.

Tilgner D.J. (1962)
Flavour analysis.
(Letter to editor)
Food Tech., 16, (7), 8.

Usborne W.R., Sokolowski J.H., Breidenstein B.C., Doane B.B.
Hatfield E.E. and Garrigus U.S. (1961)
Effect of sex on organoleptic qualities of young lamb.
J. Anim Sci., 20, 922A.

Vandore J.F. (1967)
MSc Thesis
Dept. of Agric., University of Edinburgh.

Vesely J.A. (1973)
Fatty acids and steroids affecting flavour and aroma from ram, cryptorchid and wether lambs.
Can. J. Anim. Sci., 53, 673-678.

Vesely J.A. and Hironaka R. (1976)
Feedlot performance, carcass traits and flavour of lambs fed all concentrates or hay and concentrates diets.
Can. J. Anim. Soc., 65, 51-57.

Von Sydow E and Akesson C.(1977)
Correlating instrumental and sensory flavour data.
Sensory Properties of Foods. (Eds.) Birch G.G., Brennan J.G., and Parker K.J.
Applied Science Publishers. 113-127

Wasserman A.E. and Talley F. (1968)
Organoleptic identification of roasted beef, veal, lamb and pork as affected by fat.
J. Food Sci., 33, 219-223.

Wasserman A.E.and Spinell A.M. (1972)
Effect of some water soluble components on aroma of heated adipose tissue.
J. Agric. Food Chem., 20, 171-174.

Wasserman A.E. (1979)
Chemical basis for meat flavour: a review.
J. Food Sci., 44, 6-11.

Watson R.H.J. (1981)
The importance of colour in food psychology
in Counsell J.N. (Ed.) Natural Colours for Food and Other Uses
Applied Science Publishers. 27-38.

Wenlock W.R., Buss D.H., Derry B.J. and Dixon E.J. (1980)
Household food wastage in Britain.
Brit. J. Nut., 43, 53-70.

Wheeler J.L., Park R.J., Spurway R.A. and Ford A.L. (1974)
Variation in the effects of forage rape on meat flavour in sheep.
J. Agric. Sci., Camb 1974, 83, 569-571

Wilson L.L., Ziegler J.H., Rugh M.C., Merritt T.L., Simpson M.J. and Kreuzberger F.L. (1970)
Comparison of live, slaughter and carcass characteristics of rams, induced crytorchids and wethers.
J.Anim.Sci., 31, 455-461.

Winger R.J. and Pope C.G. (1981)
Selection and training of panellists for sensory evaluation of meat flavours.
J. Food Tech., 16, 661-669.

Woodhams P.R., Kirton A.H. and Jury K.E. (1966a)
Palatability characteristics of cross breds as related to individual Southdown ires, slaughter age and carcass fatness.
N.Z.J. Agr. Res., 9, 28-85.

Woodhams P.R. Kirton A.H., and Jury K.E. (1966b)
Palatability characteristics of crossbred lambs as related to individual Southdown sires, slaughter age and carcass.
N.Z.J. Agric. Res., 9, 268-275.

Woodward W.A. and Schucany W.R. (1977)
Combination of a preference pattern with the triangle test.
Biometrics, 33, 31-39.

Wong E., Johnson C.B. and Nixon N.L. (1975) The contribution of 4-methyloctanoic (hircinoic) acid to mutton and goat flavour.
N.Z.J. Agric. Res., 18, 261-266.

Wright D.E., Payne E., and Kirton A.H. (1974)
Polyunsaturated fat in young ruminants.
N.Z.J.Agric. Res., 17, 295-297

Wright D.E., Payne E., Pyle C., Aitken W.M. and Kirton A.H. (1977)
Polyunsaturated fat in young ruminants.
Anim.Feed Sci. and Tech., 2, (1), 93-100

Yule W.J. and McBride R.L. (1979)
Lupin and rape seed meals in poultry diets: effects on broiler performance and sensory evaluation of carcasses.
Brit. J. Poultry Sci., 19, 231-239.

RELATED STUDIES

Pilot Study	1975
Evaluation for on Farm Use of De-oiled Herring Silage as a Protein Feed for Growing Pigs	1976
Trials of a Diet Containing Grain Distillers Spent Wash for Pigs	1977
A Consumer Test of Bull vs Steer Beef	1978
The Evaluation of Liquid De-oiled Herring Silage in Diets for Growing Pigs : Palatability Studies	1980

QUEEN MARGARET COLLEGE

PILOT STUDY

1 The detection of possible taints in the body fat of pigs fed non-conventional dietary regimes, using triangular testing techniques of assessment.

2 The transfer of volatiles between meat samples cooked simultaneously.

A research project carried out jointly with the Edinburgh School of Agriculture.

Organised and carried out by: Mrs. D. A. PARRY

Mrs. M. P. WOODS

SCIENCE DEPARTMENT

QUEEN MARGARET COLLEGE

1975

INDEX

INTRODUCTION

The first two experiments described in this report were carried out in the form of a pilot study to assess the feasibility of carrying out research projects jointly between staff of Queen Margaret College and the Edinburgh School of Agriculture. Samples of body fat of pigs fed with differing levels of defatted herring meal were already available for use in preliminary trials. As it is already known that the use of fish meal in the feeding of pigs may cause fishy taints in the carcass it was hoped that the defatting of the herring could avoid this problem. Initial tests indicated that sample collection and storage techniques required modification. Satisfactory procedures were however devised.

Since the olfactory sense is of greater significance than taste in flavour assessment, it was decided to make a preliminary comparison of the aroma of the body fat of test and control animals. Since it has now been established that statistically significant differences exist, an appraisal of future experimental programmes is now required. Ethically, it is of importance to establish that toxicity tests have been carried out if panel members are to be required to make extensive sensory evaluations of flavour. The design of experiments will be simplified in the future since it appears that taints are not transferred from one sample to another during cooking.

SUMMARY

The effects of immediate pre-slaughter feeding of defatted herring meal on the body fat of pigs have been studied. To allow rapid detection of possible taints, triangle tests were used. Testers, mainly second and fourth year student dietitians, were asked to assess heated samples of fats from test and control animals. Results indicate that of the nine test samples, eight showed statistically significant differences from the controls. Triangle tests indicate only that differences exist: further experiments are required to determine the nature of these differences. An additional preliminary study suggested that transfer of volatiles between meat samples cooked simultaneously is unlikely to cause difficulty in taste panel evaluations.

METHODOLOGY OF THE PILOT STUDY

Preliminary Experiments

As a preliminary to the pilot study, samples of body fat from some of the test animals were used to determine the feasibility of carrying out odour description tests. Samples had been obtained from the midline of the inguinal region of the hot carcass and subsequently frozen. Objectionable taints and extremely rapid onset of rancidity made it clear that alternative sampling techniques were essential. In subsequent tests, the use of samples of body fat from cold carcasses in the brisket region of the midline were considered to be satisfactory for the purposes of sensory evaluation.

Allocation of Test and Control Samples

Samples of body fat from nine test animals and nine controls were available. Assessments were carried out during two successive weeks. On the first occasion, test samples were compared with purchased pork fat and on the second with samples from the control animals.
The key to the samples is given below:

	Test Sample	Code	Control Sample	Code
Tested 18.6.75	1	E10H	)	
	2	E16H	)	Purchased
	3	E17H	)	Pork
	4	E18H	)	Fat
	5	E19H	)	
Tested 24.6.75	5	E19H	5	E0942H
25.6.75	6	E56H	6	E9482H
	7	E57H	7	I/DE9349H
	8	E58H	8	E9351H
	9	E59H	9	E9241H

It will be noted that, in most cases, samples were allocated numbers which corresponded with original codings.

Selection and Design of Testing Techniques

At this stage, since it was necessary to establish only if there were differences between test and control samples, triangle testing techniques were used. In this type of test, testers are presented with three samples two of which are identical and the third different. Testers are required only to identify the odd sample. Ideally, to avoid subject bias, six presentations should be made for assessment : ABA, BAA, AAB, BAB, ABB and BBA, where A and B represent control and test samples respectively. Thus to assess only the first five samples, 90 sample tubes would have been required.

Because this was considered to be an unrealistic number of sample tubes to be presented to testers on a single occasion and, that at the high level of dietary supplementation used, there was a reasonable possibility that taints might be present, the experiment was restricted to presentations ABA, BAA and AAB where A and B, as previously indicated, were control and test samples respectively. A specimen of the forms used to record testers' assessments is included as an Appendix to this report.

It will be noted that it is necessary to ensure that sample B is presented as P, A or Z on an equal number of occasions and that the allocation of the three presentations of each test sample is made at random. The design of the experiment in each of the testing sessions is indicated overleaf:

SAMPLES	PRESENTATION	SAMPLE TUBE CODE	STATION & BEAKER NO
1 and 5	AAB	PAZ	14
	ABA	PAZ	5
	BAA	PAZ	2
2 and 6	BAA	PAZ	13
	ABA	PAZ	12
	AAB	PAZ	7
3 and 7	ABA	PAZ	3
	BAA	PAZ	15
	AAB	PAZ	6
4 and 8	ABA	PAZ	11
	AAB	PAZ	8
	BAA	PAZ	1
5 and 9	AAB	PAZ	9
	ABA	PAZ	10
	BAA	PAZ	4
where B = test sample A = control sample			

It will be noted that test sample 5 was used twice. This allowed the design of the experiment to be the same at each of the weekly testing sessions and some check on the accuracy of the procedure to be made.

Testing Procedures

Two sets of 45 clean, dry boiling tubes containing small samples of fat were prepared as indicated in the table above. Tubes were covered tightly with aluminium foil which was replaced as often as was required during the testing sessions. Each set of three tubes, labelled P, A and Z was transferred to a 250 ml beaker containing 100 ml water. The fifteen beakers were immersed in thermostatically controlled water baths for at least an hour prior to testing.

Testers/

Testers were requested to refrain from smoking for at least 30 minutes previous to the experiments, to avoid the use of perfumed cosmetics, to wash hands thoroughly in unperfumed detergent liquid and to assess samples in batches to avoid sensory fatigue. The importance of allowing beakers assessed by previous testers to become reheated sufficiently to allow the release of volatiles was also stressed. They were requested to assess the samples at the appropriate test station where recording forms were available. After completion, the forms were left face down at the station. Each of the assessment forms was named so that, in future experiments, it would be possible to identify those with little aptitude for such testing procedures and thus avoid their selection as testers.

It will be appreciated that almost all the testers were completely untrained and inexperienced and were not selected in any way other than being available at the time of the testing sessions. No details had been given as to the nature of any differences which might be predicted.

On the first occasion, the tests were duplicated in adjoining rooms and on the second on two different days.

RESULTS OF THE PILOT STUDY

WEEK OF	SAMPLE	CODE	NO. of TESTERS	CORRECT ANSWERS
16.6.75	1	E10H	83	40
"	2	E16H	85	35
"	3	E17H	86	55
"	4	E18H	89	42
"	5	E19H	89	51
23.6.75	5	E19H	73	50
"	6	E56H	75	34
"	7	E57H	74	42
"	8	E58H	78	40
"	9	E59H	76	45

The number of assessors for each sample is not always the same. On occasions when the tester recorded "don't know" the form was discarded since this is an invalid response. In addition, the base of some of the tubes fractured which meant that the group of three was no longer available for test purposes.

Of the 45 testers, only nine participated in both sessions. Two testers identified B correctly on all occasions and 23 made the correct identification on more than 50% of the presentations. It is of course appreciated that 33⅓% of the selections would be correct on the basis of chance alone.

Nevertheless, since the actual number of testers is large, these results suggest that some taint of body fat of the test animals is occurring.

STATISTICAL EVALUATION of RESULTS

In difference tests, the chi-square (χ^2) distribution is used to test how well an observed frequence corresponds with a theoretical distribution. In triangle tests, involving one degree of freedom, χ^2 is adjusted as follows:

$$\text{Adj.}\chi^2 = \sum \frac{(|fo - fe| - \frac{1}{2})^2}{fe}$$

In the triangle test, where two of the three samples are known to be alike and the third different, the probability is ⅓ the odd sample will be chosen alone, the value of ^{2} is calculated from the expression:

$$\chi^2 = \left[(|4X1 - 2X2| - 3)^2\right] \quad /8n$$

where X1 = number of opinions favourable to sample 1
X2 = number of opinions favourable to sample 2
n = total number of trials

If the calculated value of χ^2 exceeds the significance value at any significance level it may be concluded that observed and theoretical distributions do not agree. Tables which have been prepared* have been used to assess the significance of the tabulated results of the pilot study.

* After Roessler, Warren and Guyman, Food Research 13 503-5 published in Principles of Sensory Evaluation of Food. Amerine, Pangborn and Roessler, Academic Press, pp 526-7.

SAMPLE	CODE	No. of Testers	CORRECT ANSWERS	Minimum correct judgements to establish significant differentiation.			STATISTICAL DIFFERENCE
				P = 0.05	P = 0.01	P = 0.001	
1	E10H	83	40	37	39	42	HS
2	E16H	85	35	37	40	43	-
3	E17H	86	55	38	40	44	VHS
4	E18H	89	42	39	42	45	HS
5	E19H	89	51	39	42	45	VHS
5	E19H	73	50	33	35	38	VHS
6	E56H	75	34	34	36	39	S
7	E57H	74	42	33	36	39	VHS
8	E58H	78	40	35	37	40	VHS
9	E59H	76	45	34	36	39	VHS

DISCUSSION OF SIGNIFICANCE OF RESULTS

Of the nine test samples, only one did not differ significantly from the control. Of the remaining eight test samples, one showed a statistically significant difference from the control at the P = 0.05 level (S), two at the P = 0.01 level (HS) and five at the P = 0.001 level (VHS), where S, HS and VHS indicate significant, highly significant and very highly significant differences respectively.

In the case of sample 5, very highly significant differences were obtained on both occasions it was tested indicating that the testing procedure had been satisfactory.

Conclusions

Results achieved in this pilot study indicate that there are statistically significant differences in the aroma of the samples of body fat from the test and control animals used in this experiment.

TRANSFER OF VOLATILES DURING THE COOKING OF MEATS

Methodology

Since kippers and garlic are well recognised for their characteristic and extremely pungent aromas, it was considered that there would be a possibility that, if they were incorporated into meat samples, that contaminating volatiles might be transferred to other, and blander, meats with which they were cooked simultaneously. Minced filleted pork was used to prepare three control samples. Four test samples were also prepared as indicated. No seasoning was used in any of the preparations. Seven identical miniature loaf tins were used in the experiment. The first and second control samples were prepared in the same way but the third had aluminium foil projecting above the tin for approximately 80 mm. The test samples were prepared as indicated. No foil was used.

SAMPLE TIN	OBSERVATIONS
Control 1	No foil 100 g minced pork
" 2	No foil 100 g minced pork
" 3	With foil 100 g minced pork
Test 1	50 g minced pork + 50 g minced kipper
" 2	100 g minced kipper
" 3	100 g minced pork + 2 cloves crushed garlic
" 4	100 g minced pork + 1 clove crushed garlic

All seven samples were cooked simultaneously on the same shelf of a preheated electric oven (205°C) until each had attained an internal temperature of at least 85°C.

Test samples were then discarded.

Control samples 1 and 3 were tasted hot and cold by DAP and MPW who had scored 28/30 and 13/14 respectively in the previous experiment.

Results

It was noted that despite mincing and mixing, the three control samples differed slightly in flavour both hot and cold. In control sample 1, 1 taster (DAP) detected a very slight kipper taint when the sample was hot. This taster assessed the samples in the order 3, 2, 1, whereas the second taster (MPW) assessed the samples in reverse order. No kipper taint was detected by either taster when the samples were cold.

Both tasters detected a very slight garlic taint in control sample 3 when cold.

Conclusions

After discussion, it was concluded that the taints were so minimal that they would be unlikely to be detected by tasters unaware that garlic and kipper had been used as a source of volatiles.

Hence, in future experiments, it would appear unlikely that samples cooked simultaneously would exert significant effects on each other and that, unless some very unexpected results were to be achieved, it would seem unnecessary to make allowance for possible transfer of volatiles between samples.

Acknowledgements

The interest, assistance and co-operation of the following people is acknowledged with gratitude:

Student dietitians and other testers from Queen Margaret College
Laboratory technicians, Queen Margaret College.
Staff of the Edinburgh School of Agriculture.

Future Experimental Programme

Subject to discussions, this might involve:

1 Feeding lower levels of dietary defatted herring meal until no statistically significant differences in the body fat could be detected.

2 Using test and control samples of fats in the preparation of meat products such as meat loaves and sausages to determine if the statistically significant differences in the aroma of the fats exerted effects in a practical situation.

3 Assessing roasted meats - hot and cold - from control and test animals of comparable age and sex cooked to constant internal temperature: Joints for roasting would require to be as far as possible matched: possibly prepared by the Carcass Dissection Unit. Cuts from different areas should be assessed since accumulation of taint may differ from one area to another.

It would be of interest to know if feed conversion ratios, growth curves and food intake are comparable for test and control animals.

Since the procedures in this pilot study appear to be feasible, the use of more sophisticated apparatus would be of assistance in the future. It is also hoped that it would be possible to select and train groups of staff and students as members of taste panels in order to achieve results of satisfactory validity.

QUEEN MARGARET COLLEGE

RESEARCH PROJECT

THE SENSORY/ORGANOLEPTIC APPRAISAL OF MEATS

BY THE USE OF TRIANGLE TESTS

Of the three samples marked P, A and Z two are identical the third is different. Please identify the odd one _____

APPENDIX II

QUEEN MARGARET COLLEGE

RESEARCH PROJECT

THE SENSORY/ORGANOLPETIC APPRAISAL OF MEATS

BY THE USE OF TRIANGLE TESTS

Of the three samples marked P, A and Z two are identical the third is different.

Please identify the odd one:-.....

Indicate by a tick whether you think the odd sample differs from the identical samples

(i) greatly
(ii) moderately
(iii) slightly

Name:-.....

Testers were requested to avoid odour fatigue by making appraisals at intervals and to observe Panel Discipline procedures.

From: *Proceedings of the Torry Research Station Symposium on Fish Silage, Aberdeen, 1976.*

EVALUATION FOR ON FARM USE OF DE-OILED HERRING SILAGE AS A PROTEIN FEED FOR GROWING PIGS

G.M. Hillyer, D.G. Peers, and R. Morrison with D.A. Parry and M.P. Woods
The Edinburgh School of Agriculture and Queen Margaret College, Edinburgh.

Introduction

Balance trials have indicated that de-oiled acidified herring silage is of high nutritive value for growing pigs. The following investigations were carried out to evaluate the use of this product as a diet ingredient for growing pigs under commercial conditions. The investigations come under three general headings, namely on-farm handling and feeding methods, feeding trials with carcase classification and studies on the odour and flavour of the meat.

On farm handling and feeding methods

It is recognised that fish silage could be handled in liquid feeding systems with only minor adjustments being necessary to existing systems. Sedimentation tests have been carried out in the laboratory and a tendency towards sedimentation has been noted with the de-oiled product. This has also been noted in bulk storage on farms. Measurements taken after four weeks show a gradient in dry matter content from the bottom to the top of the liquid, but the material was readily agitated to its original homogeneous condition. When left unagitated in bulk containers for a period of four months, a distinct deterioration had occurred with obvious signs of fermentation and therefore a more limited period of storage is necessary. Given these limitations, the main 'on farm' test for quality control and diet formulation would be a dry matter determination. The use of a hydrometer for a quick dry matter determination may have possibilities.

Although liquid feeding systems for pigs would seem to present no major problems, the majority of growing and finishing pigs are fed with dry diets and therefore dry mixing and distribution feeding equipment has also been tested to determine its suitability for handling silage containing diets of 60% dry matter.

A number of feedmixers have been tested and none have proved satisfactory. Fountain and vertical type mixers in common use have proved unsuitable for mixing diets with a dry matter content of 80% or less. Propionic acid and molasses applicators have also been tested without success in their current design.

Dry feed distribution equipment has been tested but bridging and clogging of diets with diets of dry matter of 70% occurs in feed hoppers and feed dispensers. It is concluded that the incorporation of fish silage as an ingredient into dry diets would not allow satisfactory mixing and distribution with current equipment.

However, fish silage alone or with water could be distributed separately into troughs and allow the pigs to mix the diet during normal eating. This system has worked satisfactorily in the feeding trial at the College pig unit.

Feeding Trials (College pig unit)

Diets. Previous results of chemical analyses and balance trials had suggested that the dry matter of the herring silage would be slightly superior to that of white fish meal. However in the commercial situation soya bean meal rather than fish meal would normally be used as the main supplementary source of protein for growing pigs. In feeding trials therefore diets containing herring silage at a medium (MS) diet and high level (HS diet) of inclusion level, 75 kg and 150 kg per tonne dry matter (DM) respectively were compared with diets based on soya bean meal and formulated to contain equivalent levels of total lysine (diets MC and HC). The medium inclusion level was considered to be suitable for commercial use and the high inclusion level was adopted to highlight any possible effects of the fish silage on the odour and flavour of the pig fat. The composition of the diets is shown in Table 1.

Table 1

Composition of diets (kg/tonne DM)

	High lysine		Medium lysine	
	Control (HC)	Silage (HS)	Control (MC)	Silage (MS)
De-oiled herring silage	-	150	-	75
Soya bean meal	288	-	147	-
Barley meal	679	842	818	902
Limestone	5	5	6	5
Dicalcium phosphate	25	-	26	15
Vit. mineral mix *	3	3	3	3

* Coopers Nutrition Products (Product No. 10TE)

Water was added to all diets to bring them to the same dry matter content of 600 kg/tonne.

Five batches of de-oiled herring silage were received during the investigation and the variation between these batches is shown in Table 2.

Table 2

Variation in chemical composition between batches of de-oiled herring silage

	Mean	Range
Dry matter (g/kg)	221	202 - 267
pH	3.7	3.6 - 3.9
Composition of dry matter (g/kg)		
Nitrogen	105	92 - 118
Crude protein (N x 6.25)	656	575 - 738
Ether extract	42.0	29.3 - 50.0*
Ash	164	153 - 171

*Samples > 50 g ether extract/kg were rejected.

Dry matter, nitrogen and fat were the most variable and these variations necessitated reformulation of diets to maintain equal dry matter and lysine levels.

The chemical analysis of the four diets is given in Table 3.

Table 3

Chemical composition of diets (kg/tonne DM)

Diet	HC	HS	MC	MS
Digestible energy (MJ/kg DM)*	14.9	14.9	14.5	14.5
Crude protein (N x 6.25)	219	191	164	148
Lysine	12.5	12.7	8.5	8.6
Methionine & cystine*	7.0	7.5	5.7	5.9
Crude fibre	51	46	51	49
Ether extract	18	25	20	23
Ash	77	57	67	60
Calcium	10	11	10	11
Phosphorus	9	9	9	9

*Calculated values from individual ingredients.

The analysis confirmed that the diets containing herring silage to have similar levels of energy and lysine to their respective control soya diets. The lower lysine concentration in soya bean meal accounted for the higher crude protein content in the control diets HC and MC.

Rations. All pigs were fed to a time based scale. Pigs were introduced to either treatment or control diets on Mondays and fed at a level of 1.1 kg DM/pig/day. No palatability problems were experienced with any of the diets. All pigs were weighed individually on Tuesdays and Wednesdays of the same week giving a mean starting weight and then the time based feeding scale introduced on the Thursday and changed every successive Thursday regardless of pig weights. The scale is shown in Table 4.

Table 4

Feeding scale

Week on trial	Ration (kg DM/pig/day)
1	1.32
2	1.50
3	1.68
4	1.86
5	2.04
6 + weeks	2.22

Animals. In stage I of the investigation pigs were grown from 30 to 60 kg liveweight. Sixteen pigs per pen selected at random but balanced for sex (8 females and 8 castrates) were allocated to either treatment to control diets. Three replicates were completed involving 48 pigs per treatment and 192 pigs in total. The pigs were slaughtered when weighing approximately 60 kg liveweight and the results are shown in Table 5.

Table 5

	Diet					
	HS	HC	MS	MC	SE Diff	Sign
Growth rate (g/day)	683	535	690	522	±22	***
Killing out %	69.9	70.1	70.8	70.5	±0.65	NS
Probe $P_1 + P_3$ (mm)	24.1	26.0	26.2	24.2	±1.8	NS
Feed conversion efficiency (DM)	2.5	3.8	2.4	3.3	±0.36	*

Growth rates of pigs fed fish silage diets were significantly higher ($P > .001$) than either of the control diets and feed conversion efficiency of pigs fed fish silage diets was also significantly better ($P > .05$). No significant differences were found in any of the other parameters measured between killing at percentage, probe measurements or between sexes. No carcase of pigs fed either the treatment or control diets were down graded by MLC classification.

In stage II of the investigation pigs were grown from 30 to 90 kg liveweight. Twelve pigs per pen selected at random but balanced for sex (6 females and 6 castrates) were allocated to either treatment or control diets. Three replicates involving 36 pigs per treatment and 144 pigs in total were used. The results from this stage are currently being analysed.

Commercial On farm trial (Commercial unit)

A further feeding trial was carried out on a commercial unit where 320 pigs were grown from 40 to 87 kg liveweight approximately. The composition of the diet fed on a dry matter basis was 5% fish silage, 5% soya meal, 90% barley together with a vitamin and mineral mix. The mean daily liveweight gain for all pigs was 570 g/day, and the FMC cutter pig carcase classification gave results of 80% Grade A, 15% Grade B and 5% Grade C. These grading results were typical of all pigs slaughtered from this unit over the same period, but fed a conventional barley, soya diet.

Carcase quality. No pigs on any of their feeding trials have so far been down graded by MLC carcase classification.

Studies on the Odour and Flavour of the Meat

Fat and joint samples taken from the carcases of treated and control pigs are undergoing sensory evaluation.

After preliminary trials which validated the accuracy of the technique, triangle tests were used to determine if differences in the aroma of the body fat of treated and control animals could be detected. Fat from 10 animals at each of the two treatment levels was compared with fat from appropriate controls. Statistically significant differences have been detected in some of the test samples at both treatment levels, as shown in Table 6.

Table 6

Aroma of pig fat - Sensory triangular tests

	No. Pairs of samples	Total Testers (125) Comparisons sign. different	Total Testers (125) Percent of samples correctly identified	Selected Testers (44) Comparisons sign. different	Selected Testers (44) Percent of samples correctly identified
HS v HC	10	4	43	8	60
MS v MC	10	5	46	8	60

The results including all Testers showed four and five of the paired comparisons of fat aroma to be significantly different for the HS v HC and MS v MC respectively. The results from Selected Testers on their aptitude for sensory evaluation tests, found eight of both 10 paired fat comparisons to be significantly different in aroma.

It should be emphasised that, although of statistical significance, such differences neither equate with 'taint' nor imply lack of acceptability to the consumer. An ongoing programme of paired comparison (preference) tests is at present being carried out on meats (cooked under standard conditions) from carcases of treated and control animals to determine if observed differences in body fat aroma are reflected in loss of palability.

Crown copyright reserved

Trials of a Diet Containing Grain Distillers Spent Wash for Pigs

D A Parry (Unpublished 1977)

As indicated in the Annual Report of the East of Scotland College of Agriculture (1979), the author carried out palatability studies using the methodology of the Deoiled Herring Silage trials. Joints from eight pigs on the test diet were compared with eight matched controls in May 1977. Triangle tests were carried out on raw and cooked fatty tissue. Paired comparison (preference) tests were used to assess lean tissue after pork loins[1] had been roasted to an internal temperature of 87.5°C at an oven temperature of 177°C. When cold, meats were sliced, visible fat removed and slices coverted to uniform $10mm^2$ portions for presentation to tasters. Tasters were requested to indicate preference on the basis of flavour. Results of tests are indicated below.

Triangle Tests : Aroma of Raw and Cooked Fatty Tissue

Triangle Tests - Aroma of Raw Fatty Tissue

Trial	Total Responses	Correct Identifications	Statistical Significance
1	17	8	NS
2	18	10	$p<0.05$
3	18	16	NS
4	18	8	NS
5	14	14	$p<0.001$
6	15	14	$p<0.001$
7	15	5	NS
8	15	8	NS

Triangle/

[1] Lumbar vertebrae 10-13.

Triangle Tests - Aroma of Cooked Fatty Tissue

Trial	Total Responses	Correct Identifications	Statistical Significance
1	21	8	NS
2	21	6	NS
3	21	6	NS
4	22	4	NS
5	28	19	$p<0.001$
6	29	10	NS
7	30	9	NS
8	30	12	NS

Levels of statistical significance were derived from Appendix Table E (Amerine et al 1965).

Thus, whilst statistically significant differences were demonstrated in the aroma of raw fat from three of the eight test samples, in only one trial did this difference persist after cooking. In this trial (Trial 5) the number of correct identifications in the triangle tests was so high as to indicate that factors other than diet may have contributed to difference in aroma of fat from test and matched control animals. As can be seen from the results of paired comparisons, this difference in aroma was not reflected in flavour preference.

Paired Comparison (Preference) Tests : Flavour

Trial	Total Responses	Preference for Test Sample	Statistical Significance
1	17	9	NS
2	17	11	NS
3	17	10	NS
4	17	6	NS
5	23	12	NS
6	23	13	NS
7	23	17	$p<0.05$
8	23	17	$p<0.05$

Thus/

Thus except in the fourth trial, the flavour of joints from pigs on the test feeding regime was preferred to those from matched controls. Two of these preferences were of statistical significance.

Conclusions

Although only eight comparisons were made it seems likely that the inclusion of grain distillers spent wash in the diet of pigs is beneficial rather than detrimental to the flavour of roasted loins from these animals. Differences in aroma of raw fatty tissue in three animals persisted after cooking on only one occasion. Aroma differences had no influence on flavour preferences. On the basis of these trials this dietary component can be recommended for inclusion in the diet of pigs.

J. Sci. Fd Agric. 1978, **29**, 885–894

A Consumer Test of Bull *vs* Steer Beef

Peter Baron, David Lesser,[a] Ian Robertson[b], Doreen Parry,[c]
Basil Lowman,[d] Neil Scott,[d] and John Prescott[d]

[a] *Department of Agricultural Marketing, University of Newcastle upon Tyne,* [b] *Royal (Dick) School of Veterinary Studies, University of Edinburgh,* [c] *Department of Science and Diabetics, Queen Margaret College, and* [d] *East of Scotland College of Agriculture*

(*Manuscript received 6 March 1978*)

Previous studies of the acceptability of bull beef have relied on laboratory panel or mechanical tests. This study extends the assessment to a consumer panel who cooked and ate the conventionally prepared steaks in their own homes. The test showed consumers to distinguish bull steaks as leaner than steer steaks but not to discriminate on overall appearance. When it came to eating, however, bull steaks were judged to take longer to brown and be less satisfactory in terms of flavour, juiciness, tenderness and overall eating quality. On tenderness 11 % of the sample marked the bull steaks as 'well below average'. These results are sufficiently clear to suggest the need for a test of their commercial impact.

1. Introduction

There have in recent years been extensive trials and a considerable increase in the use of bull beef. The production economies associated with bull beef are now well established.[1] It is accepted that the beef from bulls is leaner and particularly acceptable to self-service sale.[2] There is also a widely held belief that bull beef is acceptable to consumers and some evidence to support this.[1] A trained taste panel, however, while finding no major differences in aroma, flavour, tenderness and juiciness, did find bull meat to be slightly less tender than steer meat, but nevertheless to be quite acceptable.[3] Similarly Rhodes[1] reported no significant differences in flavour, juiciness and overall acceptability but differences in colour and tenderness in both mechanical tests and trained panel tests with meat from young animals. The difference in tenderness has also been confirmed elsewhere.[4, 5]

Although there is some information based on trained taste panel tests, more direct information on consumer acceptability is relatively sparse. In view of this and the increasing use of bull beef, an extensive consumer test was undertaken in conjunction with an investigation on bull beef production from suckler herds which was conducted by Robertson and Lowman[6] in the East of Scotland. At the outset of the study the principal null hypothesis was that there would be no detectable differences in the acceptability of beef or its eating quality whether it came from bulls or steers. Rump and sirloin steaks were selected for testing as it was considered that consumers would be most sensitive to their acceptability in terms of both appearance and eating quality.

2. Experimental

2.1. Origin and preparation of the beef

Rump and sirloin steaks were taken from bull and steer carcasses, which were obtained from cattle supplied on four of the farms cooperating in the study described in greater detail elsewhere.[6] They were selected to represent typical carcasses from the the groups of cattle in their investigation. These were beef cross calves suckled to 8.5 to 10.5 months of age and finished for slaughter on a variety of commercial diets at 14 to 18.5 months of age. The production and rearing methods employed and carcasses produced were considered to be highly representative of local commercial suckler calf

0022-5142/78/1000-0885 $02.00 © 1978 Society of Chemical Industry

enterprises. Each bull/steer pair was selected on the same farm of origin as being fit for slaughter on the same day, and was drawn from the group of cattle in which they had been reared and finished as contemporaries. Each pair of animals was handled, transported and slaughtered together in a standard manner. They were segregated from other animals when transported and slaughtered immediately on arrival at the abattoir. The distance transported was between 5 and 20 miles. At slaughter there was no dark cutting beef. The physical characteristics of the carcasses used in the study are shown in Table 1.

Table 1. Physical characteristics of the carcasses used to provide meat for the consumer studies

Test	Sex	Breed	Age at slaughter (days)	Weight at slaughter (kg)	Killing out (%)	M.l.c. classification	Dissection data		
							Lean (%)	Fat (%)	Bone (%)
1	B	H × BG	453	1036	53.2	3 L	64.1	17.0	14.6
	S	H × BG	419	870	51.0	3 H	53.9	25.9	14.7
1A	B	H × BG	453	1010	52.1	3 L	—	—	—
	S	H × BG	517	960	53.7	3 H	—	—	—
2[a]	B	H × H cross	539	1035	54.8	2	67.3	15.0	13.8
	S	H × H cross	538	985	52.1	4	55.4	26.4	14.2
2A[a]	B	H × H cross	533	1030	53.1	3 L	—	—	—
	S	H × H cross	538	975	52.4	3 H	—	—	—
3[a]	B	ND × H cross	536	1020	55.2	3 L	—	—	—
	S	ND × H cross	544	970	52.6	3 L	—	—	—
3A[a]	B	ND × H cross	542	1020	54.2	2	72.1	9.8	15.1
	S	ND × H cross	557	995	55.1	2	65.7	16.5	15.0

[a] Sides used for the standardised cooking and panel tests.
H, Hereford; BG, Blue Grey; ND, North Devon.
—, Not available.

Conditioning and handling of the pairs of carcasses was standardised. Rump and sirloin joints were cut from the left side of each pair of carcasses at the same lapse of time after slaughter, following standard commercial cutting procedures. The bone-in joints were immediately transported to the Meat Laboratory where they were boned-out and cut into a series of 12.70 mm thick steaks numbered in relation to their anatomical location. The steaks were trimmed, if necessary, to a fat cover of approximately 15 mm but only a proportion of the steer steaks required trimming. The steaks were placed, after preparation, in polystyrene trays and covered with film in a conventional supermarket form and were promptly delivered to the homes of the consumers.

2.2. Cooking and taste panel test

Sirloin roasts of 600–1150 g weight were obtained from four of the six pairs of carcasses (identified in Table 1[a]) and were used for standardised cooking tests and panel tests for flavour. These sample joints were sealed in polythene film and stored frozen until required for testing, when they were thawed before cooking in an identical manner. They were roasted in tins at an oven temperature of 163 C to an internal temperature of 74 C. Cooking times, final weights and the weight gain of the tins (representing dripping lost from the joint) were recorded to the nearest gram. The joints were carved into slices from which visible fat was removed before being presented cold to the tasters. At least 50 untrained tasters were used for each paired comparison. Each panelist was offered a pair of samples on each occasion and asked to express a preference for flavour. The samples were presented alternately from bull and steer carcasses.

2.3. Consumer trial: design and procedure

The consumer test of the acceptability of bull and steer steaks involved the person who normally purchased meat for the household receiving the prepacks of steaks and cooking them in the home.

Each household was given a pair of steaks to assess on two successive weekends and two identified members of the household were asked to assess the eating quality of the steaks on each occasion. In the first weekend half the number of households was given steaks from the bull and the other half was given steaks from the steer; in the second weekend each household was given steaks from the alternative source. Care was taken to ensure that for each household steaks from the same anatomical location were provided each weekend.

There were three sets of consumer tests, each with 60 households. The times of the three trials were: 1 and 1A, May 1975; 2 and 2A, February 1976; 3 and 3A, May 1976. These tests have each been analysed separately but, because few differences were observed between them, combined results for the three tests are presented here.

2.4. Selection of consumer sample

For purposes of the consumer test a random selection of households was drawn from the Electoral Roll for four prescribed areas of south Edinburgh. These areas were selected to provide households of a range of socio-economic groups. A preliminary visit was made to establish their willingness to cooperate and a total of 179 households were recruited. This sample yielded 162 usable questionnaires from two people in each household; a total sample of 324 consumers commenting on the palatability of the steaks. Table 2 illustrates the socio-economic classification of households with, for comparison, the distribution for Great Britain,[7] and their age distribution. It is apparent that compared with national figures the sample over-represented categories A and B and under-represented categories C_2, D and E. Analysis of the results by socio-economic class has, however, shown that this anomaly in the characteristics of the sample did not significantly affect the results.[a]

Table 2. Sample characteristics and socio-economic classification of households

		A	B	C1	C2	D	E	Total
Sample	No.	25	40	41	32	18	6	162
	%	15.4	24.7	25.3	19.8	11.1	3.7	100.0
G.B.	%	2	10	24	32	22	9	100.0

	Age of eaters			
	16–21	21–60	Over 60	Total
No.	21	236	60	317
%	6.6	74.4	18.9	100

2.5. Consumer questionnaire

A questionnaire delivered with the steaks was to be completed by the cook and by two eaters in each household. The cooks were asked to rate the steaks on: overall appearance, leanness, colour of lean and texture of lean, and to state the characteristic used to judge overall quality and comment on the appearance of the steak. Details of cooking method and comments on cooking were sought. The two people in the household who ate the steaks were asked to complete assessments for: overall eating quality, flavour, juiciness and tenderness. In addition eaters were asked to state which characteristics they considered important as influencing overall eating quality and to comment on

[a] A least squares analysis of covariance shows a significant class effect and interaction of class and bull score. The effect is, however, slight and still leaves a substantial and significant difference between bull and steer scores. To illustrate the slight impact of these effects, weighted and raw, mean and differenced scores for tenderness can be examined.

	Weighted by G.B. proportions		Raw proportions	
Bull	2.85	difference = 0.79	2.71	difference = 0.93
Steer	3.63		3.64	

In view of this slight impact a more complex analysis separating out this effect is not included.

the steaks. A five-point rating scale was used for each characteristic and one of these is illustrated below:

For LEANNESS do you rate it as:

5	4	3	2	1
Well above average	Above average	Average	Below average	Well below average
☐	☐	☐	☐	☐

Two standards of comparison were used. Respondents were asked each week to compare the sample steaks with the steak they bought usually. In the second week they were also asked, separately, to compare the steak with that received the week before. This second method of appraisal provided a controlled comparison and counters the objection that, from courtesy, people would over-rate something received free.

3. Results

3.1. Cooking and taste panel test

The differences arising on cooking sirloin roasts from four pairs of bulls and steers are shown in Table 3. The bull and steer joints attained the same internal temperature in a similar cooking time. Dripping loss was greater from steer roasts than from bull roasts, but because of the greater volatile loss from bull joints the overall weight loss on cooking was similar for both types of beef. The joints looked similar after roasting, but on slicing the bull beef showed a less complete colour change and appeared 'less well-done'. This colour difference shows up clearly on colour photographs but no analytical measures were taken. In the panel test of the eating quality of lean slices from the roasts, the majority of tasters showed a preference for steer beef in terms of flavour. As Table 4 shows, 62% preferred the steer beef compared with 38% for the bull beef.

Table 3. Average cooking time and cooking losses from bull and steer sirloin joints on roasting

	Bull	Steer
Uncooked weight (g)	962.7	912.1
Cooking time (min)	133.0	133.0
Cooking losses (% of wet weight)		
Volatile	22.2	16.8
Dripping	4.6	10.1
Total	26.8	26.9

Table 4. Numbers of tasters preferring bull or steer roast sirloin for flavour

Bull steer pair	Bull	Steer
2	42	91
2A	30	40
3	27	39
3A	29	38
Overall	128 (38%)	208 (62%)

$\chi^2 = 18.6^{***}$ on 4 df (if pair 2 excluded, $\chi^2 = 6.2^{*}$ on 3 df).

3.2. Consumer test

The principal results of the consumer test are presented as follows: (a) in terms of rating scores compared with steaks usually purchased, (b) as 'look-back' comparisons between the steak received in the second week and that received in the first, and (c) as differenced scores for the rating of bull and steer steaks when compared with the steak usually purchased.

(a) Comparison with steaks usually purchased

Appearance

When consumers were asked what characteristic they considered most important in assessing the quality of steaks on the basis of appearance, no predominant feature emerged (Table 5). Examination of the scores for appearance (Table 6) showed no major differences between bull and steer

Table 5. Characteristics influencing assessment of quality on appearance (percentage of 162 consumers considering the characteristic 'most important')

	Bull	Steer
Leanness	37.5	30.6
Colour of lean	25.0	33.1
Texture of lean	36.9	35.0
Other	0.6	1.2

Table 6. Mean scores for appearance (comparison with the steak usually purchased — 162 consumers)

	Bull	Steer
Overall appearance	3.42	3.44
Leanness	3.52	3.01
Colour of lean	3.35	3.46
Texture of lean	3.49	3.51
Comments on appearance		
None	77.2	78.4
Favourable	13.6	12.3
Unfavourable	9.3	9.3

steaks except for leanness. For this characteristic, 54% of respondents rated the bull steaks as above or well above average compared to only 30% for the steer steaks. There were few comments on the steaks' appearance, each type receiving roughly the same number of favourable and unfavourable (Table 6). There were no real differences in the method of cooking or additives applied to the bull or steer samples (Table 7). Half the steaks were grilled and half were fried. A quarter of the samples were cooked with added salt and pepper, 10% with other additives and 65% without any additives. Comments on cooking, however, revealed a greater number of unfavourable comments for the bull sample as regards the time taken to cook. A typical unfavourable comment was 'it takes longer to brown'.

Eating quality

In judging overall eating quality, flavour and tenderness are almost equally mentioned as the most important characteristic (Table 8). In assessing eating quality consumers did differentiate between

Table 7. Cooking methods and comments (percentage of 162 cooks)

	Bull	Steer
Cooking method		
Grilled	50.6	50.6
Fried	49.4	49.4
Additives		
Salt and pepper	25.9	24.1
Other	10.5	9.9
None	63.6	66.0
Comments on cooking		
None	67.3	71.0
Favourable	18.5	20.5
Unfavourable	14.2	2.5

Table 8. Characteristics influencing assessment of eating quality (percentage of 324 consumers considering the characteristic 'most important')

	Bull	Steer
Flavour	43.7	46.2
Juiciness	12.3	9.2
Tenderness	42.4	43.5
Other	1.6	0.6

bull and steer steaks (Table 9). Steer steaks had higher mean scores for overall eating quality, flavour, juiciness, and tenderness. The difference was especially marked for tenderness. In addition, more people made unfavourable comments on bull than steer steaks (Table 9) with regard to overall eating quality.

(*b*) *Look-back comparison with first week's steak*

The results of the look-back comparison are presented in Table 10 and have been analysed statistically using the appropriate chi-square test. This tests the null hypothesis that the scores are the same when steer is received in the second week as when bull is received in the second week. If the null hypothesis is accepted respondents detect no real differences in the two types

Table 9. Mean scores for eating quality compared with steaks usually purchased (324 consumers)

	Bull	Steer
Overall eating quality	2.90	3·63
Flavour	2.98	3.38
Juiciness	3.01	3.50
Tenderness	2.71	3.64
Comments on eating quality		
None	68.0	67.1
Favourable	11.3	23.5
Unfavourable	20.7	9.1

Table 10. Mean look-back scores for appearance and eating quality*

	Bull *vs* (Steer) [a]	Steer *vs* (Bull) [b]	χ^2	Level of significance
Overall appearance	3.39	3.35	1.81	N.s.
Leanness	3.60	2.74	32.89	***
Colour of lean	3.20	3.38	2.69	N.s.
Texture of lean	3.04	3.20	3.52	N.s.
Overall eating quality	2.63	3.79	87.40	***
Flavour	2.93	3.60	37.25	***
Juiciness	2.92	3.51	31.85	***
Tenderness	2.33	3.67	104.46	***

[a] Steer given first weekend: bull second.
[b] Bull given first weekend: steer second.
* The χ^2 statistic is calculated from the number using each score value.

Examination of the results shows that for appearance, significant differences are obtained only for leanness, with bull being considered leaner than steer. Sixty-one per cent of respondents claimed that bull was leaner than last week's compared with 21 % looking back from steer. Eating assessment revealed significant differences for all characteristics; steer was always rated above the bull sample. All the look-back scales are consistent with the results presented in the previous comparisons (Tables 6 and 9).

(c) Difference scores

The score given by each individual for bull and steer in comparison with the steak usually purchased were used to calculate difference scores. The difference calculated was bull score minus steer score. Since the basic scores have an approximately normal distribution these can be used in a paired 't' test of whether or not there is a significant difference between samples. The possible range of values is from '−4' when steer is rated much superior to bull (steer given score 5, bull scored 1) to '+4' when bull is rated much superior to steer (steer scored 1, bull scored 5). The results in Table 11 confirm the evidence presented earlier. Recipients detected bull as leaner than steer and on eating rated steer higher than bull, especially for leanness. In general, difference scores were a half point higher for leanness, and declined by a half point for flavour and juiciness, a whole point for tenderness, and three-quarters of a point for overall eating preference.

Table 11. Differenced scores in comparison with usual meat purchased (bull minus steer)

Appearance	Leanness	Colour of lean	Texture of lean	Overall appearance
Mean diff.	0.45	−0.12	−0.07	−0.2
Paired 't'	5.59***	1.58 N.s.	0.88 N.s.	0.25 N.s.
Eating quality	**Flavour**	**Juiciness**	**Tenderness**	**Overall eating quality**
Mean diff.	−0.41	−0.495	−0.93	−0.735
Paired 't'	5.86***	7.07***	12.70***	11.31***

4. Discussion

Somewhat surprisingly in the light of previous research and popular beliefs, it appears that these consumers with the samples of meat provided could distinguish bull beef from steer beef. The

results for appearance suggested that purchasers were unable to detect any marked differences in steaks compared with their normal purchases except for leanness. Fifty-four per cent of the respondents rated bull steaks above average for leanness compared with only 30% for steer steaks. In their 'look-back' comparison of steaks, 61% of consumers assessed their bull steaks leaner than their previous steer steaks whereas only 21% assessed their steer steaks as leaner than their previous bull steaks.

Both the laboratory cooking trials and the household panel suggested that the bull beef took longer to brown. Unfortunately no analytical measures are available to quantify these observations. This difference in cooking characteristics has not been reported in previous comparisons of bull and steer beef. Flavour and tenderness were identified by the consumers as the most important characteristics influencing their assessment of eating quality. They detected differences between bull and steer steaks for flavour, tenderness and juiciness, rating steers superior to bulls, especially in terms of tenderness. In overall eating quality 91% of the consumers rated the steer steaks average or above average whereas 61% considered the bull steaks were average or above (Table 12). In terms of tenderness 37% rated the bull steaks 'below average' and 11% 'well above average'; the comparable proportions for steer steaks were only 11 and 2%. Steer steaks were also generally preferred to bull steaks for eating quality when consumers were 'looking back' on the steak they had eaten the previous week. Relatively few of the consumers made specific comments on eating quality. Of these, however, 21% made unfavourable comments on the bull steaks compared with 9% unfavourable comments on the steer steaks.

Table 12. Proportion of consumers rating bull and steer steaks above or below average for appearance and eating quality (%)

	Bull			Steer		
	Average or above	Below average	Well below average	Average or above	Below average	Well below average
Overall appearance	94	5	1	89	10	1
Leanness	90	9	1	73	24	3
Overall eating quality	61	35	3	91	8	1
Flavour	67	28	5	83	14	3
Juiciness	67	30	3	85	13	2
Tenderness	52	37	11	87	11	2

All of the tests (a, b and c) showed that consumers preferred steer steaks to bull as did the tasters in a panel test. This suggests that the differences detected in other studies by trained panelists are sufficient to be noticed by non-selected testers in a conventional household environment.

No attempt was made to trim the steaks to identical fatness. This, together with the impossibility of adjusting intramuscular fat, meant that fatness was confounded with the bull–steer comparison. We know of no studies where bull and steer beef of equal fatness has been compared. Because this was a test based on currently representative production and sales methods the comparison remains valid. Moreover, fatness has little or no relationship to tenderness and contributes only slightly to juiciness and flavour.[9] The differences detected in eating quality are, therefore, attributed mainly to castration. The slower browning of bull beef might, however, be explained by its lack of fat since fat has a greater heat-carrying capacity.

In order to examine whether a particular animal was responsible for the lower ratings of bull steaks, the look-back scores for tenderness were used to detect the individuals who most clearly distinguished bull from steer beef. Tenderness is used since it is widely accepted as the most important measure of eating quality. Thirty-seven individuals marked the steer steaks as much better (score of 5) than bull when eaten last and 23 individuals marked bull as much worse than steer when eaten last. These 60 individuals who discriminated most markedly were isolated in order to check which bull they had received.

Table 13 presents the data on tenderness relating to these 60 individuals. Only one bull and one steer was used in each of the six weekend tests thus, the data relate to a restricted sample of six bulls and six steers. Significant differences among bulls at the 10% level are detected and examination of Table 13 suggests that these arise principally because of the much worse performance of bull number 3 and a much better performance of bull number 3A. It would seem, therefore, that there may be a difference between individual bulls but this does not invalidate the conclusions that in general bulls are detected as different from steers, since even in the case of the least distinguished bull (3A), 8% of consumers rate it as 'much worse' than steer.

Table 13. Number of 'much worse' scores for each bull

	Tenderness compared to last week	
Bull	No. with 'much worse' score	No. with other score
1	6	35
1A	7	42
2	14	44
2A	12	40
3	17	43
3A	4	49

$\chi^2 = 10.6^*$ on df.

Deductions about the commercial relevance of the difference between bulls and steers must be limited by the nature of the experiment, which set out to establish whether bull steaks could be distinguished from steer steaks and it has been clearly shown that they can be. A commercial test might have used a finer gradation of scales so as to distinguish the really extreme cases of consumer dislike, as had been used elsewhere for bacon.[8] Within the limitations of the 5-point scale used, however, it is worth noting that on appearance, only 1% or less scored the bull 'well above average', and on the eating scales only 3 or 4%, but in the case of tenderness 11% marked the bull extremely unfavourably. Examination of the mean scores for bull and steer samples show that steer scored, overall, slightly above the average value of 3 and bulls slightly below the average of 3. The indication is, therefore, that the sample provided to households was, in the case of steer, slightly better than their usual beef and in the case of bull, slightly poorer than their usual beef. Consequently, any comparison between weeks, as presented in Tables 10 and 11 would, while providing a more precise test, tend to overemphasise the difference found by consumers in comparison to their usual steaks. Even so, the difference scores show only 1% or less of the sample marking the bull at the extremely unfavourable score of -4.

A final point to be discussed arises because consumers did distinguish clearly the leaner nature of bull steaks. It is possible that experience or instruction has taught housewives that fat is desirable in better quality steaks. They might, therefore, have been induced to give 'prestige' answers to questions on overall appearance and eating quality. In this experiment it was not possible to resolve such a problem unequivocally. There are reasons for suggesting the prestige effect to be slight. Firstly, housewives recorded difficulty in browning bull steaks; a feature very unlikely to be associated by them with leanness. Secondly, the eating scales were also completed for the most part by husbands, presumably without training, who still scored steers preferentially. Finally, even in the 48 households scoring both types of steak the same for leanness, eating scores revealed a similar ability to distinguish bull from steer.

5. Conclusions

In general the experiment has supported the evidence that consumers can detect the greater leanness of bull steaks but this does not seem to have been sufficient to provide a preference on overall appearance. When it came to cooking and eating, the results indicated that consumers did detect a difference between bull and steer steaks and found that bull took longer to brown and that it was less satisfactory in terms of flavour, juiciness and tenderness which led to a lower score on overall eating quality. Only on tenderness, however, did the below average judgements reach sizeable proportions and this test, since it did not use a scale with fine gradations, is likely to have over-emphasised the proportion of consumers who would reject the bull steak. The fact that untrained consumers can detect bull from steer is, nevertheless, an important finding and suggests the need for further commercial appraisal, preferably involving consumers in a real choice between the meat types over several weeks.

Acknowledgements

The authors would like to acknowledge the assistance of farmers who provided cattle, and slaughter houses in the East of Scotland area who provided facilities for handling and selection of carcasses, in particular the assistance of Mr J. C. Fraser of the Carcass Evaluation Unit, Bush Estate, Penicuik, Midlothian who prepared steaks for the consumer tests. The financial support of the Scottish Agricultural Development Council for this work is gratefully acknowledged.

References

1. Rhodes, D. N. The quality of meat from male and non-male animals. *Meat Production from Entire Male Animals* (Rhodes, D. N. Ed.), Churchill, London, 1969, p. 189.
2. *Fmrs Wkly* 1976, 15 October., p. 60.
3. Trower, S. J. *N. Z. J. Agric. Res.* 1965, **8**, 921.
4. Field, R. A.; Nelms, G. E.; Shoonover, C. D. *J. Anim. Sci.* 1966, **25**, 360.
5. Joseph, R. L.; Connolly, J. *Irish J. Agric. Res.* 1974, **13**, 307.
6. Robertson, I. S.; Lowman, B. G. *Anim. Prodn* 1978, **27**, 191.
7. Monk, D. *Social Grading on the National Readership Survey* Research Services Ltd, London, 1976.
8. Lesser, D.; Baron, P. J.; Robb, J. D. *J. Sci. Fd Agric.* 1977, **28**, 1120.
9. Preston, T. R.; Willis, M. *Intensive Beef Production* Pergamon, Oxford, 1970.

J. Sci. Food Agric. 1982, **33**, 11–15

The Evaluation of Liquid De-oiled Herring Silage in Diets for Growing Pigs: Palatability Studies

Doreen A. Parry, G. Michael Hillyer and James C. Fraser

Queen Margaret College, Clerwood Terrace, Edinburgh EH12 8TS and Edinburgh School of Agriculture, Edinburgh EH9 3JG

(*Manuscript received 8 May 1980*)

Triangle tests were used to detect possible differences in the aroma of subcutaneous fat of growing pigs receiving de-oiled herring offal silage as a protein supplement at high (HS) and medium (MS) levels, i.e. at levels of 150 and 75 kg t^{-1} dry matter (DM), when compared with samples from matched controls. Tests were carried out on samples from pigs slaughtered at 60 and 90 kg. Statistically significant differences in the aroma of subcutaneous fat were detected at both feeding levels (HS and MS) and at each slaughter weight (60 and 90 kg). Correct identification of the odd samples was 60% in pigs slaughtered at 60 kg at both feeding levels. At 90 kg slaughter weight, there were 56 and 76% correct identifications at HS and MS levels, respectively. Paired Comparison (Preference) tests were used to determine if such differences were associated with flavour defects and consequent loss of palatability. Roasted pork joints and pork rusk loaves from HS and MS pigs were compared with matched controls (HC and MC). As a result of these tests it was concluded that detectable differences from controls in aroma of subcutaneous fat in samples from HS and MS pigs were not associated with alteration in consumer preference for cooked meats. There was no evidence of taint in meats from either HS or MS groups.

1. Introduction

Feeding trials by Smith and Adamson[1] suggest that levels of *non* de-oiled herring offal silage in the diet of growing pigs must be restricted to avoid production of carcass taint. To avoid this problem, Opstvedt[2] suggested that de-oiled herring offal silage may be more suitable for use as a dietary protein supplement for pigs. Sensory evaluations were therefore carried out to determine if carcasses of pigs given high (HS) and medium (MS) levels of protein supplement [150 and 75 kg t^{-1} dry matter (DM)] in the experiment reported by Hillyer *et al.*[3] showed either evidence of taint or produced meats with flavour defects likely to reduce consumer acceptance.

Because fat and fat-soluble precursors influence meat flavour, and fat may act as a solvent for substances likely to exert a modifying effect on flavour, examination of fat depots may provide warning of possible flavour defects. Triangle testing techniques had been shown in a pilot study to be effective in demonstrating differences between the aroma of subcutaneous fats from the upper thoracic mid-line region. Triangle tests indicate only that differences do or do not exist between samples. To avoid positional bias, ideally six presentations should be made. In this experiment where there was a reasonable possibility that differences existed, to avoid sensory fatigue, presentations were restricted to three.

Paired Comparison (Preference) tests were used for flavour evaluation. Were statistically significant differences in aroma to be established it would be possible to determine if these differences were associated with loss of palatability as a result.

For both Triangle tests and Paired Comparisons (Preference) statistical tables were used or the Chi-square statistic calculated[4] to evaluate results.

0022–5142 82 0100–0015 $02.00 © 1982 Society of Chemical Industry

2. Experimental

2.1. Aroma evaluation

2.1.1. Triangle tests

Samples of subcutaneous fat from ten HS and ten MS pigs slaughtered at 60 kg were compared with matched controls (HC and MC). To avoid positional bias HS and MS were presented as the odd samples at first, second and third positions in the same series of tests. Samples from the upper thoracic mid-line region were cut using separate scalpels and transferred to labelled boiling tubes with ground glass stoppers. Each batch of three tubes was placed in a numbered beaker immersed in a thermostatically controlled water-bath at 60 C. Testers were required to identify by aroma which tube contained the odd sample. A special area of the laboratory was used for each test. Recording forms were retained in the same area. The experiments were repeated comparing samples of subcutaneous fat from HS and MS pigs slaughtered at 90 kg with matched controls. These samples were from the same carcasses as the roasted pork joints assessed in the Paired Comparison (Preference) tests. Agreement between results of tests could therefore be determined.

2.2. Paired Comparison

2.2.1. (Preference) tests

Each coded sample was presented first on an equal number of occasions. Effects of the subjective nature of individual preference were minimised by a large number of presentations.

Samples from eight HS and eight MS pigs slaughtered at 90 kg were compared. Joints from lumbar vertebrae 10–13 were used. Previous tests had established that cross-contamination of samples during cooking does not occur. Thus, to achieve standard rates of heating each pair of samples was cooked on a trivet in tins located on the same shelf of an electric oven pre-heated to 177 C The cooking process was considered to be completed when an internal temperature of 87.5 C was attained.

To eliminate textural differences which might influence flavour preference, pork rusk loaves were prepared using a model sausage mixture derived from Gerard.[5] Cooking and internal temperatures were the same as for roasted joints. Standard covered loaf tins were used.

To allow continued uniformity in presentation during the tasting sessions both roasted joints and pork rusk loaves were sliced and portioned cold. Visible fat was removed to ensure consistency in the appearance of samples; this is recognised consumer practice and is a dietary recommendation.[6]

Table 1. Aroma of the subcutaneous fat of 60-kg pigs—Triangle tests

Test	Number of comparisons	Significance—all testers				Significance—selected testers[a]			
		NS	*	**	***	NS	*	**	***
HS vs HC	10	6	2	0	2	2	2	0	6
MS vs MC	10	5	3	0	2	2	1	0	7

[a] Testers selected on ability to identify odd sample correctly in at least 50% of presentations.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS = Not significant.

Analysis of testers' performance

	Total responses	Correct responses	Percentage correct
HS vs HC			
All testers	1805	789	43.7
Selected testers	536	323	60.3
MS vs MC			
All testers	1377	622	45.2
Selected testers	455	270	59.3

Whilst it could be argued that tasting of hot samples might have been a more sensitive procedure in detecting flavour differences, experiments in the laboratories of the authors indicate that tasters' acuity is not adversely affected when meats are tasted cold.

3. Results

3.1. Triangle tests

Results of Triangle tests to detect differences in aroma of subcutaneous fat of ten HS and ten MS pigs slaughtered at 60 kg are given in Table 1. Table 2 shows the results of the Triangle tests for the 90-kg pigs.

Table 2. Aroma of the subcutaneous fat of 90-kg pigs—Triangle tests

Test	Number of comparisons	Significance—all testers				Significance—selected testers[a]			
		NS	*	**	***	NS	*	**	***
HS *vs* HC	8	5	1	0	2	5	1	0	2
MS *vs* MC	8	2	0	2	4	2	0	1	5

[a] See footnotes to Table 1.

Analysis of testers' performance

	Total response	Correct response	Percentage correct
HS vs HC			
All testers	282	134	47.5
Selected testers	162	90	55.6
MS vs MC			
All testers	281	166	59.1
Selected testers	155	118	76.1

3.2. Paired Comparison (Preference) tests

Results of Paired Comparison (Preference) tests on roasted joints and pork/rusk loaves to detect differences in flavour are given in Table 3.

4. Discussion

Statistically significant differences in aroma of subcutaneous fat were established at both supplement levels and at both slaughter weights. The sensitivity of the Triangle testing procedure was increased by using only results of testers showing aptitude for the testing procedure in that correct judgments were given on at least 50% of presentations. Detection of difference in the 60-kg pigs increased from 4 and 5 to 8 in both HS and MS groups, respectively, when results provided only by these testers were used. For the 90-kg pigs, improvement in testers' performance was shown but this was not reflected in increased sensitivity of the Triangle testing procedure. Differences were detected in only three of eight HS *vs* HC comparisons, whereas in six of eight of the MS and MC comparisons differences were noted. From Tables 1 and 2 it can be seen that greater differences in aroma of subcutaneous fat from carcasses at both 60 and 90 kg slaughter weights were demonstrated between MS and MC groups than between HS and HC groups.

Triangle tests indicated only that detectable differences existed between tests and control samples. Hedonic ratings or descriptive analyses were not required. In the Paired Comparison (Preference)

Table 3. Paired Comparison (Preference) tests—flavour of roasted joints and pork rusk loaves from (number preferring samples from different treatments) 90-kg pigs

Comparison: HS *vs* HC	Roasted pork joints HS	HC	*n*	Significance	Pork rusk loaves HS	HC	*n*	Significance
1	20	15	35	NS	18	31	49	NS
2	23	23	46	NS	10	17	27	NS
3	29	21	50	NS	48	54	102	NS
4	34	14	48	**	53	32	85	*
5	23	16	39	NS	18	16	34	NS
6	33	23	56	NS	11	17	28	NS
7	24	26	50	NS	28	22	50	NS
8	20	11	31	NS	—	—	—	—
Comparison: MS *vs* MC	MS	MC	*n*	Significance	MS	MC	*n*	Significance
1	25	25	50	NS	19	17	36	NS
2	24	11	35	*	18	21	39	NS
3	14	18	32	NS	27	23	50	NS
4	24	16	40	NS	17	15	32	NS
5	13	26	39	NS	16	14	30	NS
6	7	24	31	**	16	11	27	NS
7	14	17	31	NS	16	13	29	NS
8	15	14	29	NS	19	10	29	NS

See footnotes to Table 1.

tests, considering HS and HC roasts from the 90-kg pigs, HS joints were preferred in six of eight presentations. In only one of these presentations, however, was the preference of statistical significance ($P < 0.01$). In MS *vs* MC roasts, of the two comparisons associated with statistically significant flavour preferences, one ($P < 0.05$) was for MS and the other ($P < 0.01$) for MC. Examination of the comparisons as a whole suggests that preferences on the basis of flavour were difficult to establish with 55% of tasters preferring HS roasts ($\chi^2 = 3.41$ NS) and 54% MC roasts ($\chi^2 = 2.18$ NS). It should be noted that in the HS *vs* HC flavour comparison of statistical significance, no difference between aroma of subcutaneous fats was detected. In the MS *vs* MC comparisons, at the $P < 0.01$ level, differences in aroma were detected by Triangle tests but on one occasion flavour preference was for MS and on the other for MC.

Flavour preferences for pork/rusk loaves were even more difficult to establish. As with roasted joints, the single flavour preference ($P < 0.05$) for the HS pork rusk loaf suggests that this particular carcass was of very superior flavour. Flavour preference for HS and MS loaves was 49.6 and 54.4%, respectively ($\chi^2 = 0.01$ NS and $\chi^2 = 1.94$ NS) but as a result of converting meats to meat loaves to reduce influence of textural differences on flavour preference, comparisons showed greater lack of sensitivity than in roasted samples.

Hence, in this series of experiments it is reasonable to conclude that detectable differences in aroma of the subcutaneous fat samples did not affect flavour preference in roasted pork joints or pork/rusk loaves. Further tests would be required to establish whether substances responsible for detectable differences in aroma of the subcutaneous fat decompose or volatilise during the cooking process. Alternatively, lipids of the subcutaneous tissue of the upper thoracic mid-line region may not be sufficiently comparable to those of intramuscular fat of muscles from lumber vertebrae 10–13 to predict flavour defects in cooked meats.

Acknowledgements

The authors thank the laboratory technicians of Queen Margaret College, in particular Lesley Mungall, for their active participation in these investigations.

References

1. Smith, P.; Adamson, A. H. An evaluation of liquefied fish as protein source for fattening pigs. *Proc. Br. Soc. Amin. Prod.* 1976, **22**. Part 1, 16.
2. Opsvedt, J. Fish taints in eggs and poultry. *Nutrition Conference for Feed Manufacturers* (Swan, H.; Lewis, D., Eds). Churchill Livingston, London, 1971, p. 70.
3. Hillyer, G. M.; Peers, D. G.; MacAndrew, A. The evaluation of liquefied de-oiled herring silage in diets for growing pigs: Feeding trials. *J. Sci. Food Agric.* 1982, **33**, 6–10.
4. Amerine, M. A.; Pangborn, R. N.; Roessler, E. B. *Principles of the Sensory Evaluation of Food* Academic Press, London, 1965, Appendix Tables D and E, pp. 442–443, 525–527.
5. Gerard, F. *Sausage and Small Goods Production* Leonard Hill, London, 1969, pp. 116–117.
6. *Eating for Health* Department of Health and Social Security, 1978, p. 79.